Immunity and other defenses in pea aphids, Acyrthosiphon pisum.

Nicole M Gerardo, Boran Altincicek, Caroline Anselme, Hagop Atamian, Seth M Barribeau, Martin de Vos, Elizabeth Duncan, Jay D Evans, Toni Gabaldon, Murad Ghanim, et al.

Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*

N.M. Gerardo[1,§], B. Altincicek[2], C. Anselme[3,4], H. Atamian[5], S.M. Barribeau[1], M. de Vos[6], E.J. Duncan[7], J.D. Evans[8], T. Gabaldón[9], M. Ghanim[10], A. Heddi[3], I. Kaloshian[5], A. Latorre[11], A. Moya[11], A. Nakabachi[12], B.J. Parker[1], V. Pérez-Brocal[3,11], M. Pignatelli[11], Y.Rahbé[3], J. S. Ramsey[6], C. Spragg[1], J., Tamames[11], D. Tamarit[11], C. Tamborindeguy[13,14], C. Vincent-Monegat[3], A. Vilcinskas[2]

[1]Department of Biology, Emory University, O. Wayne Rollins Research Center, 1510 E. Clifton Rd. N.E., Atlanta, GA, 30322, USA.

[2]Interdisciplinary Research Center, Institute of Phytopathology and Applied Zoology, Justus-Liebig-University of Giessen, Heinrich-Buff-Ring 26-32, D-35392 Giessen, Germany.

[3]Université de Lyon, INRA, INSA-Lyon, IFR41 BioEnvironnement et Santé, UMR203 BF2I, Biologie Fonctionnelle Insectes et Interactions, Bat. Louis-Pasteur 20 ave. Albert-Einstein, F-69621 Villeurbanne, France.

[4] UMR Interactions Biotiques et Santé Végétale, INRA 1301-CNRS 6243-Université de Nice-Sophia Antipolis, 400 routes des Chappe, F-06903 Sophia-Antipolis cedex, France.

[5]Department of Nematology, Graduate Program in Genetics, Genomics and Bioinformatics, University of California, 900 University Ave., Riverside, CA, 92521, USA.

[6]Boyce Thompson Institute for Plant Research, Ithaca, New York, 14853, USA.

[7]Genetics Otago and The Laboratory for Evolution and Development, Department of Biochemistry, University of Otago, Box 56, Dunedin 9054, New Zealand

[8]USDA-ARS Bee Research Lab, BARC-East Bldg 476 Beltsville, MD 20705 USA.

[9]Bioinformatics and Genomics Programme. Centre for Genomic Regulation (CRG). Doctor Aiguader, 88. 08003 Barcelona, Spain.

[10]Department of Entomology, The Volcani Center, Bet Dagan 50250, Israel,

[11]Instituto Cavanilles de Biodiversidad y Biología Evolutiva, Universitat de València, Avenida Blasco Ibañez 13, 46071 València (Spain), CIBER en Epidemiología y Salud Pública (CIBEResp) and Centro Superior de Investigación en Salud Pública (CSISP), Conselleria de Sanidad (Generalitat Valenciana), Avenida de Cataluña, 21. 46020 València (Spain)

[12]Advanced Science Institute. RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.

[13]Plant Pathology and Plant-Microbe Biology Department, Cornell University, Tower Road, Ithaca, NY, 14853, USA.

[14]Department of Entomology, Texas A&M, College Station, TX 77843-2475, USA.

[§]Corresponding author

Email addresses:

NMG: nicole.gerardo@emory.edu

BA: boran.altincicek@agrar.uni-giessen.de

CA: caroline.anselme@sophia.inra.fr

HA: hagop.atamian@ucr.edu

SB: seth.barribeau@gmail.com

MD: md285@cornell.edu

EJD: elizabeth.duncan@otago.ac.nz

JE: jay.evans@ars.usda.gov

TG: tgabaldon@crg.es

MG: ghanim@agri.gov.il

AH: abdelaziz.heddi@insa-lyon.fr

IK: isgouhi.kaloshian@ucr.edu

AL: amparo.latorre@uv.es

AM: andres.moya@uv.es

AN: bachi@riken.jp

BJP: bparke4@emory.edu

VP: vicente.perez-brocal@uv.es

MP: miguel.pignatelli@uv.es

YR: yvan.rahbe@jouy.inra.fr

JSR: jsr47@cornell.edu

CS: cspragg@emory.edu

JT: javier.tamames@uv.es

DT: datachu@alumni.uv.es

CT: ctamborindeguy@ag.tamu.edu

CVM: carole.monegat@insa-lyon.fr

AV: Andreas.Vilcinskas@agrar.uni-giessen.de

Abstract

Background

Recent genomic analyses of arthropod defense mechanisms suggest conservation of key elements underlying responses to pathogens, parasites and stresses. At the center of pathogen-induced immune response are signaling pathways triggered by the recognition of fungal, bacterial and viral signatures. These pathways result in the production of response molecules, such as antimicrobial peptides and lysozymes, which degrade or destroy invaders. Using the recently sequenced genome of the pea aphid (*Acyrthosiphon pisum*), we conducted the first extensive annotation of the immune and stress gene repertoire of a hemipterous insect, which is phylogenetically distantly-related to previously characterized insects models.

Results

Strikingly, pea aphids appear to be missing genes present in insect genomes characterized to date and thought critical for recognition, signaling and killing of microbes. In line with results of gene annotation, experimental analyses designed to characterize immune response through the isolation of RNA transcripts and proteins from immune-challenged pea aphids uncovered few immune-related products. Gene expression studies, however, indicated some expression of immune and stress-related genes.

Conclusions

The absence of genes suspected to be essential for the insect immune response suggests that the traditional view of insect immunity may not be as broadly applicable as once thought. The limitations of the aphid immune system may be representative of

a broad range of insects, or may be aphid specific. We suggest that several aspects of the aphid life style, such as their association with microbial symbionts, could facilitate survival without strong immune protection.

Background

Aphids face numerous environmental challenges, including infection by diverse pathogens and parasites. These pressures include parasitoid wasps, which consume their hosts as they develop inside, and a variety of viral, bacterial and fungal pathogens. Both parasitoid wasp and fungal pathogens cause significant decline of natural aphid populations [1, 2], and have been suggested as potential agents for biocontrol of these agriculturally destructive pests. While facing such challenges, aphids also cope with predators and abiotic stresses, such as extreme temperature fluctuations. Thus, like most insects, aphids must attempt to survive in a harsh, complex environment.

Insects have a number of defense mechanisms. First, many insects, including aphids, behaviorally avoid predators, pathogens, and environmental stressors [3-6]. When stressors cannot be avoided, insects have a protective cuticle and gut pH inhospitable to many foreign organisms. If these barriers fail, immunological defense mechanisms recognize the invader, triggering a signaling cascade and response. While insects do not have adaptive, antigen-based responses typical of vertebrates, insects do have innate immune responses, which include clotting, phagocytosis, encapsulation, and production of antimicrobial substances [7, 8]. Phagocytosis and encapsulation, are referred to as cellular responses as they are mediated by blood cells [9]. Reponses vary depending on the invader, with antimicrobial peptides being

central to combating microbes and encapsulation being central to combating larger invaders, such as parasitoids. Until recently, it was presumed that insects were limited to these non-specific innate immune responses and had no specific immunity (*e.g.,* the antigen based immune response of humans). There is, however increasing evidence for the ability of insects to mount specific immune responses [10].

Here we focus on the identification of aphid genes that are known to play a role in the recognition and degradation of microbial pathogens in other insects, as these are the invertebrate defense processes that are best understood. In the fruit fly *Drosophila melanogaster,* recognition of an invasive microbe leads to signal production via four pathways (Toll, IMD, JNK, and JAK/STAT) [11]. Each pathway is activated in response to particular pathogens [12]. Signaling triggers the production of a multitude of effectors, including, most notably, antimicrobial peptides (AMPs). Insect AMPs may be 1000-fold induced in microbe-challenged insects compared to basal levels. In insect genomes annotated to date, these pathways appear well conserved, with most of the key components found across flies (*Drosophila* spp.), mosquitoes (*Aedes aegypti*, *Anopheles gambiae*), bees (*Apis mellifera*) and beetles (*Tribolium castaneum*) [13-17].

Because aphids and other insects face diverse challenges, we propose models for several genes critical to other elements of insect stress responses. These include genes encoding heat shock proteins (HSPs), which are synthesized in almost all living organisms when exposed to high temperatures or stress [18]. We also suggest models for genes involved in the synthesis of the alarm pheromone (*E*)-β farnesene (EBF), which aphids release in the presence of predators [19]. While there are undoubtedly many other genes involved in stress and immunological responses, our selection of

genes for exploration provides a broad survey of the known insect immune and stress repertoire and will serve as a basis for future exploration of more specific responses.

The pea aphid genome provides novel insights into arthropod immunity for two reasons. First, most of our understanding of insect immune and stress responses comes from holometabolous insects, the group of insects with complete metamorphisis, such as flies, butterflies, beetles and bees. The genome of the hemimetabolous pea aphid, *Acyrthosiphon pisum*, may thus provide novel insight into immunity and defense in more basal, non-holometabolous insects, which have incomplete metamorphisis. Second, aphids are unique amongst the arthropods sequenced to date in that they are intimately dependent on both obligate and facultative bacterial symbionts for their survival. The aphid symbiont community includes *Buchnera aphidicola*, obligate and intracellular Gram-negative bacteria that have the ability to synthesize required amino acids not readily available in the aphid diet. Beyond this obligate symbiosis, aphids frequently host one of more other Gram-negative bacterial symbionts, including most notably *Hamiltonella defensa*, *Serratia symbiotica* and *Regiella insecticola* [20, 21]. Unlike *Buchnera*, which is present in all aphids and is thus considered a *primary* symbiont, these bacteria are considered to be facultative, *secondary* symbionts, because their presence varies within an aphid species [22]. Secondary symbiotic bacteria have been shown to influence several aspects of aphid ecology, including heat tolerance and resistance to parasites and pathogens [23-26]. Specifically, both *H. defensa* and *S. symbiotica* confer protection against parasitoid wasp development [27, 28], and *R. insecticola* decreases *A. pisum* mortality after exposure to the fungal pathogen *Pandora neoaphidis* [29]. These are some of the best-studied examples of symbiont-conferred protection [30].

Aphids thus provide an excellent opportunity to study the immune system of an organism that is dependent on microbial symbionts but is hampered by parasites and pathogens. Despite this, little work has been done to characterize the aphid immune response. Altincicek et al. (2008a) found that compared to other insects, stabbing a pea aphid with bacteria elicits reduced lysozyme-like (muramidase) activity, and no detectable activity against live bacteria in hemolymph assays. Furthermore, suppression subtraction hybridization (SSH) of bacterial-challenged aphids uncovered no antimicrobial peptides and few genes of known immune function [31]. These results are surprising given that similar studies in other insects demonstrate that antimicrobial peptide production and upregulation of immune-related genes is a common feature of the insect immune response that can be captured in functional assays such as SSH [32-35]. This suggests that aphids have a significantly reduced or altered immune repertoire.

Using the recently sequenced genome of the pea aphid clone LSR1, in this study, we take two approaches to study immunity and stress in pea aphids. First, we assay presence/absence of a subset of known immune and stress-related genes. Second, we combine functional assays targeting the production of RNA and proteins to gain insight into how pea aphids respond to various challenges. Overall, our results suggest that pea aphids are missing many genes central to immune function in other insects, and that, although pea aphids do mount some response to challenges, the overall immune-response of pea aphids is more limited than that of other insects studied to date.

Results and Discussion

Overview of Annotation

We focused our manual annotation efforts on a subset of genes involved in the innate, humoral immune response contributing to recognition, signaling and response to bacteria and fungi in arthropods. We also manually annotated some genes involved in more general stress responses (*e.g.,* heat shock proteins). All annotations are based on the recently completed sequencing of pea aphid clone LSR1 [36]. All genes manually annotated, as well as those genes that we found to be missing in the pea aphid genome, are listed in Supplemental Table S1. Also in this table, BLAST-based searches revealed that another aphid, *Myzus persicae* (green peach aphid), has putative homologs for many immune and stress related genes identified in the pea aphid.

Annotation of Microbial Recognition Genes

PGRPs. Upon microbial invasion, *Drosophila* utilize several pathogen recognition receptors (PRRs) to detect pathogen-specific molecular patterns (*e.g.,* cell-surface motifs) [37]. PRRs include peptidoglycan receptor proteins (PGRPs), which recognize peptidoglycans present in cell walls of Gram-positive and Gram-negative bacteria. PGRP-based recognition activates both the Toll and IMD/JNK pathways. PGRPs are highly conserved, with mammals and insect PGRPs sharing a 160 amino acid domain [38, 39]. Thus, it is surprising that pea aphids, in contrast to all other sequenced insects, appear to have no PGRPs. One other sequenced arthropod, the crustacean *Daphia pulex,* is also missing PGRPs [40].

GNBPs. GNBPs (Gram-Negative Binding Proteins, a historical misnomer) are thought to detect Gram-positive bacteria [41]. GNBPs and PGRPs are suspected to

form a complex. GNBPs then hydrolyze Gram-positive peptidoglycans into small fragments, which are detected by PGRPs [41, 42]. Aphids have two *GNBP* paralogs, *GNBP1* and *GNBP2* (see Figure S1a in Additional data file 1). Because GNBPs are thought to form a complex with PGRPs, the presence of GNBPs without PGRPs in aphids, as well as in the crustacean *D. pulex* [40], calls into question whether GNBPs play a role in bacterial detection in these organisms. Some GNBPs and similar proteins are known to function in fungal recognition [42], which may be the primary function of these molecules in aphids.

Lectins. Lectins are a diverse group of sugar binding proteins. Many lectins function in insect immune recognition by binding to polysaccharide chains on the surface of pathogens [43]. *Drosophila* c-type lectins also appear to facilitate encapsulation of parasitoid invaders, by marking surfaces for hemocyte recruitment [44]. Aphids have five *c-type lectin* paralogs.

Galectins are another widely-distributed group of lectins [45]. In mosquitoes, *galectins* are upregulated in response to both bacterial and malaria parasite infection [46, 47]. Insect galectins are thought to be involved in either pathogen recognition, via recognition of β-galactoside, or in phagocytosis [45]. Aphids have two *galectin* paralogs.

Class C scavenger receptors. Scavenger receptors exhibit broad affinity towards both Gram-positive and Gram-negative bacteria, but not yeast [48]. Pathogen recognition by class C scavenger receptors in *Drosophila* facilitates phagocytosis, and natural genetic variation of *Drosophila* scavenger receptors is correlated with variation in the ability to suppress bacterial infection [49]. While *D. melanogaster* has four class C scavenger receptor homologs, *A. gambiae* and *A. mellifera* have only one. Pea aphids appear to have no class C scavenger receptors.

The Nimrod superfamily and Dscam. Several members of the Nimrod superfamily appear to function as receptors in phagocytosis and bacterial-binding [50, 51]. Such insect genes include *eater* and *nimrod*. Many of these genes are characterized by a specific EGF-repeat (Epidermal Growth Factor-repeat), and are duplicated in the genomes of *D. melanogaster*, *T. castaneum* and *A. mellifera* [52]. We were unable to identify any EGF-motif genes in the pea aphid genome.

Complex alternative splicing of Dscam (Down Syndrome Cell Adhesion Molecule) generates diverse surface receptors sometimes employed in arthropod innate immune defenses [53-55]. Though we did not manually annotate this complex gene as a part of this initial aphid immune gene project, we did identify multiple predicted proteins sequences in the aphid genome with strong similiarity to Dscam in other insects (GenBank: XP_001951010, XP_001949262, XP_001945921, XP_001951684, XP_001942542). Further investigations will be necessary to determine the activity and hypervariability of these genes and their transcripts in aphids.

Annotation of Signaling Pathways

The toll signaling pathway. The toll pathway is a signaling cascade involved in both development and innate immunity. In *Drosophila*, deletion of many of the component genes leads to increased susceptibility to many Gram-positive bacteria and fungal pathogens [11], and some Gram-negative bacteria and viruses [12]. In addition, upregulation of many components of the toll pathway is observed following parasitoid wasp invasion [56]. The toll pathway appears to be intact in pea aphids. We found convincing matches for genes encoding the extracellular cytokine spätzle, the transmembrane receptor Toll, the tube and MyD88 adaptors, the kinase pelle, the

inhibitor molecule cactus (a homolog of IkB), cactin, Pellino, Traf, and the transactivator dorsal (Figure 1). The latter two genes are duplicated.

As in other insects, there are several gene families associated with the toll pathway that are represented in aphids. First, aphids seem to have multiple *spätzles* that segregate with *Drosophila spätzles* 1,2,3,4 and 6 in phylogenetic analyses (see Figure S1b in Additional data file 1). Second, aphids also have a suite of serine proteases and serine protease inhibitors (serpins). Though we did not manually annotate serine proteases and serine protease inhibitors (serpins) as a part of this initial aphid immune gene project, we did identify multiple predicted protein sequences in the aphid genome with strong similarity to serine proteases and serpins in other insects. In insects, these molecules function in digestion, embryonic development and defense responses towards both microbial and parasitoid wasp invaders [57-59]. In the absence of microbial challenge, the serpin necrotic prevents activation of the toll pathway, but upon immunological challenge, the toll pathway is triggered by a cascade of serine proteases, including persephone, which is thought to be specific to fungal challenge [41]. Though it is not clear which of the many aphid serine proteases is homologous to persephone, it is likely that pea aphids have serine proteases capable of triggering the Toll pathway. Finally, aphids also have multiple genes encoding Toll receptors, which function as transmembrane receptors in both mammals and insects. While nine single-copy Toll genes have been identified in *D. melanogaster* (*Toll1* to *Toll9*), it seems that pea aphids, like other insects, lack some of these genes, but have multiple copies of others (see Figure S1c in Additional data file 1). In other organisms, some, but not all Tolls serve a role in immune function, while others function in developmental processes [60-62]. For aphids, it is not yet clear what role each Toll serves.

The JAK/STAT signaling pathway. Like the toll pathway, in *Drosophila*, the JAK/STAT (Janus Kinase / Signal Transducers and Activators of Transcription) pathway is involved in both development and immunity. The JAK/STAT pathway is the least understood of the core insect immune pathways. JAK/STAT pathway induction appears to lead to overproliferation of hemocytes, upregulation of thiolester-containing proteins (TEPs), and an antiviral response [63]. Changes in gene expression following parasitoid wasp invasion of *Drosophila* larvae suggest a role for the JAK/STAT pathway in parasitoid response [56]. Pea aphids have homologs of all core JAK/STAT genes, including genes encoding for the cytokine receptor domeless, JAK tyrosine kinase (a.k.a Hopscotch), and the STAT92E transcription factor (Figure 1). *STAT92E* appears to be duplicated. No homologs were found for *upd* (unpaired), considered a key ligand in *Drosophila* JAK/STAT induction. This ligand is also missing in other insects (e.g *A. mellifera*) [14].

IMD and JNK signaling pathways. Surprisingly, pea aphids appear to be missing many crucial components of the IMD (immunodeficiency) signaling pathway. This pathway is critical for fighting Gram-negative bacteria in *Drosophila* [11, 64], and IMD pathway member knockouts influence susceptibility to some Gram-postive bacteria and fungi as well [12]. IMD-associated genes missing in pea aphids include *PGRPs* (see above), *IMD*, *dFADD*, *Dredd* and *Relish* (*Rel*) (Figure 1). In contrast, conserved one to one orthologs of these same genes are found across *Drosophila, Apis, Aedes, Anopheles* and *Tribolium* [13]. Cursory BLAST-based searches for these genes in other arthropods, suggest that some may be missing (Figure 2). Pea aphids do have homologs for a few pathway members (*TAB*, *TAK*, *kenny*, *Iap2* and *IRD5*) (Figure 1).

While missing IMD-associated genes, pea aphids have plausible orthologs for

most components of the JNK pathway (Figure 1). In *Drosophila*, the JNK pathway regulates many developmental processes, as well as wound healing [65], and has been proposed to play a role in antimicrobial peptide gene expression and cellular immune responses [11, 66]. Genes present include *hep*, *basket*, and *JRA*. Searchers for homologs to the *Drosophila kayak (kay)* gene found an apparently similar transcription factor encoding gene in the *A. pisum* genome [GenBank: XP_001949014], but this match was largely restricted to the leucine zipper region, and failed tests of reciprocity.

The absence of IMD but presence of JNK in pea aphids is surprising as, in *Drosophila*, the IMD signaling pathway leads to activation of components of the JNK signaling pathway [11]. Specifically, when TAK, a protein kinase of the IMD pathway, is activated, it triggers the JNK pathway. Whether TAK can be activated without the rest of the IMD pathway is unknown. An alternative IMD-independent activation of JNK, via the inducer Eiger [67], has been proposed in *Drosophila* [66]. As Eiger is present in the pea aphid, this mode of activation may serve a critical role in any aphid JNK-based immune response.

Annotation of Recognition Genes

Antimicrobial peptides. Introduction of microbes into most insects leads to the production of antimicrobial peptides (AMPs) by the fat body, an insect immune-response tissue, and occasionally by hemocytes and other tissues [68-71]. These peptides are secreted into the hemolymph, where they exhibit a broad range of activities against fungi and bacteria. The mechanisms of AMP action are poorly understood, but at least in some cases (*e.g.,* drosomycin in *Drosophila*), AMPs destroy invading microbes by disrupting microbial cell membranes, leading to cell lysis [71].

Antimicrobial peptides are diverse and ubiquitous. They tend to be small molecules (<30 kDa) specialized at attacking particular microbial classes (*i.e.,* Gram-positive bacteria, fungi, *etc*.) [68, 69]. While some antimicrobial peptides are found in only a single insect group (*e.g.,* metchnikowin is found only in *Drosophila*), others are widely dispersed across eukaryotes (*e.g.,* defensins are present in fungi, plants and animals). Genomics, coupled with proteomics, has revealed that all sequenced insects, and many other insects, have multiple types of antimicrobial peptides (Figure 2). Pea aphids, surprisingly, are missing many of the antimicrobial peptides common to other insects. For example, while all insect genomes annotated thus far have genes encoding for defensins [13], homology-based searches, phylogenetic-based analyses, transcriptomics (see below), and proteomics (see below) failed to find any signatures of defensins in the pea aphid genome. The presence of defensins in the human louse *Pediculus humanus* (Figure 2), and in the ancient apterygote insect, the fire brat *Thermobia domestica* [34], suggests that defensins have been lost during aphid evolution.

Extensive searches for genes encoding for insect cecropins, drosocin (and other proline-rich arthropod AMPs), diptericin (and other glycine-rich AMPs), drosomycin, metchnikowin, formicin, moricin, spingerin, gomesin, tachyplesin, polyphemusin, andropin, gambicin, and virescein also revealed no hits. Weak hits were found for genes that encode for two antimicrobial peptides in other invertebrates: megourin [UniProtKB: P83417], originally isolated from another aphid species, the vetch aphid *Megoura viciae* (Bulet et al., unpublished) and penaeidin [UniProtKB: P81058], originally isolated from the shrimp *Penaeus vannamei*. The putative pea aphid *megourin* (scaffold EQ11086, positions 45752 – 45892), however, is highly diverged from that of *M. viciae* (31% identity) and, as compared to its *M.*

viciae counterparts, seems to have a shorter C-terminal region containing a stop-codon (see Figure S2 in Additional data file 1). Using three different primer pairs, we were unable to amplify products of this putative *Megourin* from cDNA generated for expression analyses (see below). The highly divergent *Penaeidin* [GenBank: ACYPI37769] (see Figure S2in Additional data file 1) also did not amplify from cDNA.

We found six *Thaumatin* homologs in the *A. pisum* genome that show overall sequence and predicted structure similarities to plant thaumatins (Figure 3a,b). Thaumatin-like proteins are disulfide-bridged polypeptides of about 200 residues. Some thaumatins possess antifungal activity in plant tissues after infection [72]. Recently, a thaumatin found in the beetle *Tribolium castaneum* was shown to inhibit spore germination of filamentous fungi *Beauveria bassiana* and *Fusarium culmorum* [32]. Phylogenetic analyses revealed that *A. pisum* thaumatins form a monophyletic group closely related to beetle thaumatins (Figure 3c). Since thaumatin-like genes are conspicuously absent from the genomes of *Drosophila*, *Apis*, *Anopheles, Pediculus* and *Ixodes* (Figure 2), our findings indicate that thaumatins may represent ancient defense molecules that have been lost in several insect species, or have been independently acquired in aphids and beetles. The monophyly of aphid and beetle thaumatins provides no indication of an origin of novel acquisition (Figure 3c).

Lysozymes. Lysozymes represent a family of enzymes that degrade bacterial cell walls by hydrolyzing the 1,4-beta-linkages between *N*-acetyl-D-glucosamine and *N*-acetylmuramic acid in peptidoglycan heteropolymers [73]. They are ubiquitously distributed among living organisms and are believed to be essential for defense against bacterial infection. Lysozymes are classified into several types (*i.e.,* c (chicken), g (goose), i (invertebrate), plant, bacteria and phage types). C-type

lysozymes are the most common for metazoa, being found in all vertebrates examined thus far and many invertebrates, including all the previously sequenced insects. For example, *D. melanogaster* and *A. gambiae* have at least seven and nine loci for c-type lysozymes, respectively [74, 75]. Insects also have i-type homologs, but their bacteriolytic activities are unclear [76].

Unlike other insects sequenced thus far, similarity searches demonstrated that *A. pisum* lacks genes for c-type lysozymes. The analysis further verified that the genome also lacks genes for g-type, plant-type, and phage-type lysozymes. Only three genes for i-type homologs were detected in the genome (see Figure S1d in Additional data file 1). One of them, *Lys1*, is highly expressed in the bacteriocyte [77]. Two others, *Lys2* and *Lys3*, are located adjacent to *Lys1*.

Notably, two genes that appear to have been transferred from bacterial genomes to the *A. pisum* genome encode bacteriolytic enzymes [36]. One is for a chimeric protein that consists of a eukaryotic carboxypeptidase and a bacterial lysozyme. The other (*AmiD*) encodes *N*-acetylmuramoyl-L-alanine amidase, which is not a true lysozyme (1,4-beta-N-acetylmuramidase) but similarly degrades bacterial cell walls. While some of these bacteriolotyic-related genes are highly expressed in the bacteriocyte, and lysozymes appear to be upregulated in response to some challenges (see gene expression study, below), assays of bacterioltyic activity of hemolymph from immune-challenged aphids suggest that aphid hemolymph has weak to no lysozyme-like activity [31]. Further studies will determine the role of these gene products.

Chitinases. Chitinases are enzymes that degrade chitin (a long-chain polymer of *N*-acetyl-D-glucosamine), hydrolyzing 1,4-beta-linkages between *N*-acetyl-D-glucosamines. Chitinases and lysozymes represent a superfamily of hydrolases, and

their catalytic activities are similar. Indeed, some chitinases show lysozyme activity and vice versa [73]. In insects, chitinases are used to degrade the chitin in the exoskeleton and peritrophic membrane during molting, and some are suspected to have antifungal activity, as fungal cell walls also consists of chitin [78]. Similarity searches followed by phylogenetic analyses demonstrated that the genome of *A. pisum* encodes seven genes for putative chitinase-like proteins [79]. Further studies are required to determine the biochemical properties and substrate specificity of these chitinase-like proteins.

TEPs and Tots. Some thiolester-containing proteins (TEPs) can covalently attach to pathogens and parasites in order to 'mark' them for phagocytosis [80]. Like other insects, aphids have multiple *TEP* paralogs. Both are homologous to *TEPIII* (see Figure S1e in Additional data file 1). Homologs of *TepI, TepII* and *TepIV* were not found. In contrast, no *Turandot (Tot)* genes, which encode small peptides induced by severe stress and septic injury in *Drosophila* [81-83], have been found in aphids or in other insects other than *Drosophila* spp.. Both TEPs and Tots are thought to be regulated by the JAK/STAT pathway.

Prophenoloxidase (ProPO). Phenoloxidase-mediated melanin formation characteristically accompanies wound clotting, phagocytosis and encapsulation of pathogens and parasites [84]. In insects, the inactive enzyme prophenoloxidase (ProPO) is activated by serine proteases to yield phenoloxidase [85]. Aphids appear to have two prophenoloxidase homologs (*ProPO1*, *ProPO2*) (see Figure S1f in Additional data file 1), which are homologous to *D. melanogaster Diphenol oxidase A3* (Flybase: CG2952).

Nitric oxide synthase. Production of nitric oxide is mediated by the enzyme nitric oxide synthase (NOS). Nitric oxide is a highly unstable free radical gas that has

been shown to be toxic to both parasites and pathogens. In insects, *Nos* is upregulated after both parasite and Gram-negative bacterial infection [86, 87]. Like other insects, pea aphids have one *Nos* homolog.

Heat shock proteins. Though called heat shock proteins, HSPs are produced in response to a range of stresses in both eukaryotic and prokaryotic organisms [18]. They serve as chaperones, facilitating protein-folding and stabilization, and as proteases, mediating the degradation of damaged proteins. HSPs may also serve as signaling proteins during immune responses [18, 88]. In many insects, including aphids, *HSP*s have been shown to be upregulated after septic injury and microbial infection [31, 89-92]. We identified 15 *HSP*s of varying molecular weight in pea aphids (see Figure S1g in Additional data file 1).

Gluthione-S-tranferases (GSTs). GSTs are a diverse class of enzymes that detoxify stress-causing agents, including toxic oxygen free radical species. GSTs are upregulated in some arthropods upon oxidative stress [93] and microbial challenge [89, 94]. Pea aphids have at least 18 genes encoding GSTs and many other detoxification enzymes that likely play a role in stress responses [95]. Ramsey et al. (2009) identified many of the genes encoding detoxification enzymes in *A. pisum* and in *Myzus persicae*.

Alarm pheromone production. In response to predators, aphids release an alarm pheromone that causes neighboring aphids to become more mobile and to produce more winged than unwinged offspring [19, 96]. These winged offspring have the ability to disperse to enemy-free space. While many insects produce a suite of chemicals that constitute an alarm signal, the aphid alarm pheromone is dominated by a single compound, (*E*)-β farnesene (EBF) [97]. While the genes underlying alarm pheromone production have not been fully characterized, we have identified a

Farnesyl diphosphate synthase (FPPS) and an *Isoprenyl diphosphate synthase (IPPS),* which may underlie alarm pheromone production [98].

Functional Assays

Gene expression. We utilized real-time quantitative PCR to conduct a preliminary investigation of the expression of 23 recognition, signaling and response genes in aphids subjected to a number of infection and stress treatments, (see Supplementary Materials and Table S2 in Additional data file 1). While future studies with more biological replicates will be necessary to fully survey gene regulation in the face of stress and infection, this initial survey indicates that aphids do express these genes under both control and infection/stress conditions (see Tables S4 and S5 in Additional data file 1). This suggests that these genes are functional even in the absence of many other missing immune-related genes.

One expression pattern seen in this initial survey is of particular note. Unlike other insect immune expression studies, we found no strong upregulation of antimicrobial peptides, which frequently exhibit 10-fold or greater up regulation in the face of infection. For example, while Altincicek et al. (2008b) observed 20-fold upregulation of *Thaumatins* in tribolium beetles after stabbing with lipopolysaccaride endotoxin derived from *E. coli*, we saw modest upregulation (approximately 2-fold) of only one *Thaumatin* (*Thm2*) after stabbing aphids (Supplemental Table S5). Furthermore, despite the fact that they are known to suppress fungal germination in beetles, the *Thaumatin* homologs were not upregulated after fungal infection at the time point included in this study, and were only approximately 2-fold upregulated at two additional time points and in a follow-up fungal infection experiment (data not shown) [32]. The role of thaumatins in fighting microbial infections, however, should

not be discounted, as they may function in the absence of significant upregulation (*i.e.,* they may be constitutively expressed).

Exploration of ESTs from infected and uninfected aphids. In the first of two EST-based experiments, we compared a cDNA library synthesized from the guts of *A.pisum* that had been fed a Gram-negative pathogen, *Dickeya dadantii* [99], to a cDNA library synthesized from uninfected guts. Strikingly, no standard immune-related genes, such as antimicrobial peptides, were identified in the infected sample. The main functional classes differentially expressed were the "*biopolymer metabolism*" class, many members of which were down-regulated in infected guts, and "*transport*" or "*establishment of localization*" classes, whose genes were upregulated in infected guts (Supplemental Table S6). The *"immune response"* class, in contrast, was only represented by five genes. Four of these five genes were in the uninfected library, while only one, a *leucyl-aminopeptidase*, was identified from the infected library; the immune function of leucyl-aminopeptidases is not well understood. Moreover, the "response to stress/external stimulus/biotic stimulus" classes were not overrepresented in the infected gut library.

In a separate experiment, to further identify aphid immune-relevant genes, we utilized SSH to compare cDNA from *E. coli*-infected aphids and cDNA from unchallenged aphids. To obtain genes expressed at different phases of the immune response, three RNA samples were extracted 3, 6 and 12 hours after *E. coli* infection and mixed prior to cDNA synthesis.

Among the 480 expressed sequence tags (ESTs) that were sequenced from the subtracted library [GenBank: GD185911 to GD186390], we found some genes with similarity to proteases and protease inhibitors but few other immune-related proteins. Interestingly, SSH-based EST analysis failed to identify any pathogen recognition

receptors (PRRs), such as PGRPs or GNBPs, or any antimicrobial peptides (Supplemental Table S7). It is noteworthy that this aphid experiment was conducted in parallel to a similar *Sitophilus* weevil experiment, where many immune-related genes (more than 18% of ESTs) were identified, including antibacterial peptides and PRRs [35]. This suggests that the paucity of immune genes identified in *A. pisum* is not a technical issue but may be a specific feature of aphids [31]. In addition, dot blot analysis demonstrated that only a few genes (less than 5%) were differentially expressed between *E.coli*-stabbed and unstabbed aphids. These findings indicate that, in contrast to other insects, either aphids respond only weakly to challenge with Gram-negative bacteria or aphid genes and pathways directed against these bacteria are expressed only constitutively.

High Performance Liquid Chromotography. HPLC peptide analyses targeting production of small peptides (*e.g.,* antimicrobial peptides) were run on hemolymph samples from pea aphids challenged by three microorganisms: *E. coli* (Gram-negative bacteria), *Micrococcus luteus* (Gram-positive bacteria) and *Aspergillus fumigatus* (fungi). Profiles were compared between control, infected and sterile-stabbed aphids at 6, 12 and 18 hours after challenge. When identified, the production of small peptides was maximal at 18 hours. In *E. coli*-treated samples, no upregulation could be identified (Figure 4a), in *M. luteus*-treated samples, there was modest upregulation (data not shown), and in *A. fumigatus*-treated samples, there was a significant response, though few peaks (Figure 4b). In contrast, a response profile to *E.coli* from another obligate symbiotic insect (the weevil, *Sitophilus oryzae*) exhibited at least five well-distinguishable upregulated peaks (Figure 4c). Response being restricted to Gram-postive bacteria and fungi is consistent with previous identification of megourin, an antimicrobial peptide in the aphid *Megoura viciae*, which appears to

have activity against Gram-positive bacteria and fungi, but not against Gram-negative bacteria (Bulet, unpublished). Because so few distinguishable peaks were present in the aphid samples, we did not choose to identify the associated products, but overall the presence of few inducible peptides suggests a peculiar scarcity of antimicrobial peptides in aphids.

Conclusions

Aphids are one of only a few genomic models for hemimetabolous insects, yet until recently, virtually nothing was known about aphid immune and stress response systems. Here, by coupling gene annotation with functional assays, we see evidence that aphids have some defense systems common to other arthropods (*e.g.,* the Toll and Jak/STAT signaling pathways, heat shock proteins, prophenoloxidase). Surprisingly, however, several of the genes thought central to arthropod innate immunity are missing in aphids (*e.g.,* PGRPs, the IMD signaling pathway, defensins, c-type lysozymes). This calls into question the generality of the current model of insect immunity, and it remains to be determined how aphids protect themselves from the diverse pathogens and parasites that they face.

The fact that we cannot find aphid homologs to many insect immune genes could be a consequence of the large evolutionary distance between aphids and the taxa (in most cases, flies, mosquitoes and bees) from which these genes are known (*i.e.,* the split between the ancestors of aphids and these taxa occurred approximately 350 million years ago [100]), making it challenging to find divergent genes via homology-based searches, even when using highly sensitive methods as done here. Though we cannot preclude this possibility in all cases, in some cases, similar

homology-based methods are able to recover homologs in even more distantly-related taxa. For example, querying genome databases with *Drosophila* genes via BLAST, recovers putative homologs of PGRPs and defensins in *Pediculus humanus* (human body louse) and in *Ixodes scapularis* (deer tick) (Figure 2). The divergence time between *Drosophila* and these taxa is equal to or greater than that between *Drosophila* and aphids. Moreover, for some cases, we could identify genomic regions similar to functional genes in other species, but these regions contain large insertions or stop codons (*e.g.,* the putative antimicrobial peptide Megourin), indicating they are the result of pseudogenization.

One potential explanation for the lack of known immune-related genes in pea aphids is that aphids mount an alternative, but equal, immune-response. Our functional analyses, as well as those of Altincicek et al. [31], found little evidence for an alternative response. In EST and HPLC analyses, few novel ESTs or peptide signals were recovered from immune-challenge aphids relative to their unchallenged controls. It should be noted, however, that these challenges were primarily limited to exposure to *E. coli* bacteria. When testing for expression of a few immune genes in response to a wider array of challenges, we do see some evidence of an aphid immune and stress response. Future expression studies, including large-scale transcriptional and proteomic studies, will extend this work and allow for more comprehensive characterization of the full complementation of aphid immune responses.

While we have focused mainly on the humoral component of the innate immune response, it is interesting to note that there is some evidence that the cellular component of pea aphids' innate immune response may also be different to that seen in other insects. While many insects encapsulate parasitoid wasp larvae, smothering them to death with hemocytes (insect immune cells), aphids appear not to have this

layer of protection [101, 102]. Aphids, however, appear to recruit some hemocytes to parasitoid eggs, suggesting that cellular immunity may play an alternative, though possibly more limited role [101]. Better insights into the capacity of the aphid immune system will require further investigation of both the humoral and cellular components of aphid immunity.

The lack of genomic and molecular data regarding immune systems of aphid relatives makes it difficult to establish whether the pea aphid immune system is unique. There are, however, a number of aspects of aphid ecology that could facilitate ecological success without a strong immune defense. Altincicek *et al.* [31] proposed three hypotheses to explain the apparent lack of antimicrobial defenses. First, they suggested that contrary to *Drosophila,* whose natural environment consists of decaying fruit that is colonized by microbes, aphids exploit phloem sap, which is usually sterile [103]. Thus, the risk of encountering pathogens while feeding is limited. This assumption, however, is only partly true. While probing plants, aphids are capable of acquiring pathogenic bacteria from the surface of their host plants' leaves [104], and aphids become host to a diverse assemblage of bacteria and fungi under stressful conditions [105], some of which are pathogenic (Gerardo, unpublished data). Furthermore, *Sitophilus* weevils, which when challenged with *E. coli* significantly upregulate immune genes [35], spend their entire larval and nymph stages within sterile cereal grains, indicating that a sterile diet is not likely to explain the absence of antibacterial defenses in aphids.

Altincicek *et al.* [31] also suggest that aphids may invest in terminal reproduction in response to an immune challenge, rather than in a costly immune response. In their study, stabbed aphids produced significantly more offspring than untreated aphids within 24 hours of injury. Such an increase in reproduction upon

challenge is not uncommon for invertebrates. *Biomphalaria* snails [106] [107], *Acheta* crickets [108], *Daphnia* waterfleas [109], and *Drosophila* flies [110] have all been shown to increase their investment in reproduction in response to infection. Yet, *Drosophila* still mount a complex immune response. Furthermore, aphids do not increase their reproductive effort in the face of all immune challenges: fungal infection reduces the number of offspring *A. pisum* produce within 24 hours of inoculation, and response to stabbing with bacteria seems to be specific to the aphid genotype and to the location of the stab [111, 112]. Therefore, though aphids have the capacity to reproduce many offspring prior to succumbing to some pathogens, it seems that immune competence would still provide increased fitness.

Even without increased reproduction following infection, the prolific reproductive capacity of aphids suggests these insects, in general, may invest most resources towards rapid, early onset reproduction rather than towards fewer, though better-protected offspring (a.k.a., in terms of classical ecological theory, aphids may be r-selected rather k-selection organisms [113]). Recent theory of the evolution of immunity suggests that such organisms may specifically invest less in costly immune responses [114, 115]. Many characteristics of aphids, including their rapid generation time, short life span and small body size all fit a model of r-selection [116]. *Drosophila* spp., however, also exhibit many of these characteristics and still invest in a strong defense repertoire.

The third hypothesis proposed by Altincicek et al. (2008a) concerning the evolution and maintenance of aphid defense relies on the presence of secondary symbionts that can be found extracellularly in aphids [117]. *A. pisum* is protected against fungal pathogens by one of these secondary symbionts, *Regiella insecticola* [29] and also against the parasitoid wasp *Aphidius ervi* by another secondary

symbiont, *Hamiltonella defensa* [27]. Such symbiont-mediated host protection may explain why aphids have a reduced (or specialized) antimicrobial defense. This hypothesis seems plausible with regard to the cost of immune gene expression versus the benefit of protection by the secondary endosymbionts. However, it does not explain how the secondary endosymbionts (as Gram-negative bacteria), often present in aphid hemolymph, are themselves perceived and controlled by the aphid immune system. Thus, it is challenging to say whether the presence of secondary symbionts is a cause or a consequence of reduced antimicrobial activity.

Potentially, all of these forces could shape the evolution of aphid stress and immune responses. In order to test these hypotheses (*e.g.,* reproductive investment, symbiont-mediated host protection), we need more studies characterizing the global aphid response under more conditions, and in more aphid species. Potential insight from aphid relatives with different lifestyles (*e.g.,* those not associated with secondary symbionts, or those that live in soil or other microbe-rich habitats) may be particularly helpful. More broadly, as the pea aphid is the first published genome of a hemimetabolous insect, future analyses of the immune and stress related genes of more insects in this group will facilitate the reconstruction of the evolutionary history of innate immunity and other defenses.

Materials and methods

Bioinformatic Screening of the Pea Aphid Genome

Immune and stress gene candidates from other insects (*e.g., D. melanogaster*, *A. aegypti*, *A. gambiae*, *A. mellifera*) were used to query the pea aphid genome. Most searches utilized the blastp search function to search for hits against the predicted *A.*

pisum proteome [118]. For some gene families and putative paralogs, protein sequences were aligned to sequences from other insects and outgroups using ClustalW [119]. These alignments, as well as available EST and full length cDNA sequences, served to refine aphid gene models (exon/intron boundaries, *etc.*), and to facilitate phylogenetic analyses. In addition, a comprehensive database of all available EST sequences from the green peach aphid, *Myzus persicae*, was screened using tblastn to search for potential homologs to all immune and stress genes annotated in the pea aphid.

For genes that could not be found in the proteome, we also conducted a tblastn search against all contigs and unassembled reads. Then, a final, more sensitive profile-based search was performed for those immune defense proteins that produced no hits with BLAST searches. For this analysis, insect and other species protein sequences belonging to the family of interest were retrieved from NCBI and aligned with MUSCLE [120]. A hidden Markov model for the alignment was built and calibrated using HMMER [121]. This was used to perform a profile-based search (hmmsearch) against the six-frame translated sequences of the assembled pea aphid genome and the unassembled reads. Additionally, a similar search with PFAM profiles [122] was also performed for those families encoding PFAM domains in their sequences. Whenever a significant hit was found, the genomic region was analyzed to discard the possibility that it encoded a pseudogene (presence of stop codons, absence of relevant domains, *etc.*).

Phylogenetic analyses of selected protein families were performed using their corresponding Maximum Likelihood phylogenetic trees from the pea aphid phylome [36], deposited in PhylomeDB [123]. When necessary, additional sequences were added to the original PhylomeDB alignment, realigned with MUSCLE and used to

reconstruct a Maximum Likelihood phylogenetic tree, using the JTT model as implemented in PhyML v2.4.4 [124], assuming a discrete gamma-distribution model with four rate categories and invariant sites, and estimating the gamma shape parameter and the fraction of invariant sites. Cladograms were edited using Dendrogram [125].

Exploration of ESTs from Infected and Uninfected Aphids

In the first experiment, two EST libraries (one control, one infected) were generated by standard procedures using a SMART cDNA kit (Clontech), starting from approximately 1000 dissected *A. pisum* midguts for each library. The aphids were clonal, young, reproducing asexuals, which were either fed on control diet or infected by feeding on artificial diet with the Gram-negative aphid pathogen *Dickeya dadantii* at 10^6 bacteria per mL [99]. Twenty-four hours after infection, control and treated aphids were dissected, and complete guts were transferred immediately into RNeasy solution (Qiagen). ESTs were sequenced according to procedures in Sabater-Munoz et al. (2006) [126].

In another EST-based experiment utilizing Suppression Subtractive Hybridization (SSH) and dot-blot technology, we treated aphids (clone LL01) with rifampicin as described in Rahbé *et al.* [127] to reduce symbiont load. We challenged wingless fourth-instar aposymbiotic aphids by stabbing them with needles previously dipped into a pellet of overnight cultures of *E. coli* (TOP10, Invitrogen), and then maintained them on fava plants. At three, six, and twelve hours post-treatment, we stored surviving aphids at -80°C. To identify genes that are differentially expressed in response to septic injury, we performed SSH using RNAs from immune challenged (3, 6 and 12 hours post-treatment) and untreated aposymbiotic aphids, using the SMART PCR cDNA synthesis Kit and the PCR-Select cDNA subtraction kit

(Clontech laboratories) according to the manufacturer's instructions and as described in Anselme *et al*. [35]. After transformation by electroporation, we recovered approximately 1500 colonies from LB agar plates. We plasmid extracted and sequenced 500 randomly picked colonies (NucleoSpin® Plasmid Kit, Macherey-Nagel) utilizing the sequencing center at the University of Valencia (Spain). We compared all sequences against UniProt using blastx. Immune-related gene sequences (Supplemental Table S7) were then compared to the aphid genome using blastn.

To analyze the differential expression status of each expressed sequence tag (EST) we conducted a dot-blot experiment. Briefly, we amplified 344 ESTs from the SSH library by colony PCR with nested PCR primers 1 and 2R from the PCR-Select cDNA subtraction Kit. We then spotted 10 µL from each PCR product onto two different membranes (Hybon™-N, Amersham) using a *Bio-Dot* Microfiltration System (Biorad). We hybridized membranes with radiolabeled cDNA probes generated by reverse-transcription from RNA extracted from either aposymbiotic aphids stabbed with *E. coli* or unstabbed aposymbiotic aphids. We synthesized these probes using the Super Script™ First Strand synthesis (Invitrogen) system for RT-PCR and [α-^{32}P]dCTP, and purified them using *Quick Spin* Column (Roche molecular Biochemicals). After exposing blots for up to 24 hours to a Storm PhosphorImager imaging plate (Amersham), we analyzed differential expression by comparison of band intensities between the two membranes. We, however, did not normalize the data, as we failed to see any signal from the *gapdh* gene, though the same amount of each PCR product was loaded on both membranes.

HPLC

Aphids were challenged by abdominal puncture with triple-0 needles dipped in a solution of Gram-negative bacteria (*E. coli strain Top10*), Gram-positive bacteria

(*M. luteus*) or fungal spores (*A. fumigatus*). For each microbial treatment, five hemolymph samples from 50 aphids each were collected at four times points (t = 0, 6, 12 and 18 hours).

Hemolymph was flash-extracted by centrifuging (1 min, 10,000g, 4°C) live aphids through a 1 mL pipette tip and directly into 40 μL 0.1% trifluoractetic acid (TFA) contaning 10 μL of saturated phenylthiourea (PTU) for phenoloxidase inhibition. Resulting samples were highly similar to pure hemolymph samples obtained by leg bleeding (> 95% band identity by silver-stained SDS-PAGE).

After initial collection, tips were removed and the samples were centrifuged for 5 minutes at 15,000g. Following addition of 70 μL TFA 0.1%, the supernatant sat for 1 hour at 4°C to allow for protein precipitation prior to a final 10 min centrifugation at 15,000g to recover peptides. Samples were evaporated and stored at -20°C until use in HPLC. Chromatography was performed on standard peptide C18-300Å reverse phase columns using water acetonitrile gradients [128]. For retention time (RT) standardization, PTU served as an internal standard, and samples were analysed by area-normalization to unchallenged sample peaks (RT = 14 min, preceding PTU).

Abbreviations

ALPV: aphid lethal paralysis virus; aLRT: approximate likelihood ratio test; AMP: antimicrobial peptide; Ct: comparative threshold cycle; DSCAM: Down Syndrome Cell Adhesion Molecule; EBF: E-β farnesene; EGF: epidermal growth factor; EST: expressed sequence tag; GNBP: gram-negative binding protein; HPLC: high performance liquid chromatography; HSP: heat shock protein; IMD:

immunodeficiency; JAK/STAT: janus kinase / signal transducers and activators of transcription; JTT: Jones-Taylor-Thornton; PDA: potato dextrose agar; PGRP: petidoglycan receptor protein; ProPO: prophenoloxidase; PRR: pathogen recognition receptor; PTU: phenylthiourea; SSH: suppression subtractive hybridization; RT: retention time; RQ: relative quantity; Tep: thiolester-containing protein; TFA: trifluoracetic acid; Tot: turandot.

Authors' contributions

NMG, SMB, and MG were group leaders for the project. NMG, BA, HA, SMB, MDV, EJD, JDE, AM, MG, IK, AN, BJP, MP, JSR, JT, DT, and CT designed and performed manual gene annotation. TG and SMB conducted phylogenetic analyses. BA and AV conceived of and conducted analyses of Thaumatin. SMB, NMG, CS and BJP performed experiments and analyses for the gene expression study. CA, AH, VPB, AM, and AL conceived of and conducted the SSH study, and CVM constructed the aphid gut libraries. YR conducted the HPLC study. The manuscript was prepared by NMG, SMB, CA, TG and YR with input from MDV, BA, AN, AV and AH. All authors have read and approved the final version of the manuscript.

Acknowledgements

We thank Angela Douglas, Nancy Moran, Tom Little and members of the International Aphid Genomics Consortium for insightful discussion, and Charles Godfray and two anonymous reviewers for comments that enhanced this manuscript. Cultures of *Z. occidentalis* were provided by the USDA ARS Collection of Entomopathogenic Fungal Cultures. Samples of ALPV and virus-infection protocols were provided kindly by Bryony Bonning and Liljana Georgievska. Comparison with *M. persicae* was supported by USDA grant 2005-35604-15446 to Georg Jander.

References

Figure legends

Figure 1 - Some key insect recognition, signaling and response genes are missing in the pea aphid.

Previously sequenced genomes of other insects (flies, mosquitoes, bees, beetles) have indicated that immune signaling pathways, seen here, are conserved across insects. In aphids, missing IMD pathway members (dashed lines) include those involved in recognition (PGRPs) and signaling (IMD, dFADD, Dredd, REL). Genes encoding antimicrobial peptides common in other insects, including defensins and cecropins, are also missing. In contrast, we found putative homologs for most genes central to the Toll, JNK and JAK/STAT signaling pathways.

Figure 2 - Gene families implicated in arthropod immunity suggest unique features of the pea aphid immune system.

Black indicates present (copy number is indicated, when known), white indicates absent, and gray indicates equivocal or unknown. Values for *D. melanogaster, A. gambiae, T. castanateum, A. mellifera*, and some *D. pulex* genes are based on published analyses [13, 14, 16, 17, 40]. For previously unannotated *D. pulex* genes, as well as for *I. scapularis* and *P. humanus* genes, we determined presence via cursory BLAST searches against available genome databases (wfleabase.org, vectorbase.org) using both *D. melanogaster* and *A. pisum* protein sequences as queries. Gene presence for *Ixodes* was confirmed based on previous studies [129]. Future comprehensive

annotation of the *Pedicularis* and *Ixodes* immune gene sets may reveal the presence of additional genes and lack of functionality of others.

Figure 3 - Evolutionarily conserved thaumatins are present in pea aphids and plants.

(a) The three-dimensional structure of the pea aphid thaumatin ACYPI009605 (left) was calculated using the published crystallographic structure of a sweet cherry (plant) thaumatin 2AHN_A (right) [130] and Swissmodel [131], revealing that both thaumatins are similar in structure. However, one exposed loop, indicated by a dashed circle, shows a significant difference in structure, suggesting possible adaptation to different targets. (b) Similarities are also revealed in the alignment of the pea aphid thaumatin with the plant thaumatin. A predicted signal sequence of the pea aphid thaumatin is underlined. Identical amino acids are highlighted in red. (c) Maximum likelihood phylogeny of thaumatins, indicating branches leading to nematode, plant, insect and bacteria-specific clades. Red highlights the sweet cherry thaumatin. Blue highlights the pea aphid thaumatins. * indicates approximate likelihood ratio test (aLRT) support > 80. (Api: *A. pisum*; Cac: *Catenulispora acidiphila*; Cel: *Caenorhabditis elegans*; Mtr: *Medicago truncatula*; Pav: *Prunus avium*; Tca: *Tribolium castaneum*; Tpr: *Trifolium pratense*)

Figure 4 - HPLC traces of inducible hemolymph peptides in the pea aphid compared to the rice weevil.

Representative traces (solid, red lines) are from insects 18 hours after microbial challenge; traces generated from 18 hour control insects are overlaid (dashed, black lines). Phenylthiourea (indicated as PTU) served as an internal standard. Arrows indicate peaks that are significantly upregulated (solid, red arrows) or downregulated (dashed, black arrows). (a) Profile from pea aphids challenged with *E. coli,* showing no upregulated response. (b) Profile from pea aphids challenged with the fungus *A. fumigatus*, showing some differential peaks. (c) For comparison, profile from rice weevils (*Sitophilus oryzae*) challenged with *E. coli*, showing several differentials peaks at multiple retention times.

Additional files

Additional data file 1 – Supplemental Material

A single supplementary document includes: Supplementary methods for the gene expression study; Table S1, Pea aphid immune and stress gene list; Table S2, Samples for qPCR expression study; Table S3, Primers for qPCR expression study; Table S4, Relative expression of recognition and signalling genes; Table S5, Relative expression of response genes; Table S6, Gut EST library statistics; Table S7, List of selected ESTs from the subtracted library; Figure S1, Maximum likelihood phylogenies of selected immune and stress gene families; Figure S2, Alignments of putative antimicrobial peptides megourin and penaeidin; and Figure S3, Survival curves for experimental infections associated with qPCR study.

References

1. Snyder WE, Ives AR: **Interactions between specialist and generalist natural enemies: Parasitoids, predators, and pea aphid biocontrol**. *Ecology* 2003, **84**: 91-107.
2. Hufbauer RA: **Aphid population dynamics: does resistance to parasitism influence population size?** *Ecological Entomology* 2002, **27**: 25-32.
3. Hatano E, Kunert G, Bartram S, Boland W, Gershenzon J, Weisser WW: **Do aphid colonies amplify their emission of alarm pheromone?** *Journal of Chemical Ecology* 2008, **34**: 1149-1152.
4. Tarpy DR: **Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth**. *Proceedings of the Royal Society of London Series B-Biological Sciences* 2003, **270**: 99-103.
5. Ha EM, Oh CT, Ryu JH, Bae YS, Kang SW, Jang IH, Brey PT, Lee WJ: **An antioxidant system required for host protection against gut infection in *Drosophila***. *Developmental Cell* 2005, **8**: 125-132.
6. Francke DL, Harmon JP, Harvey CT, Ives AR: **Pea aphid dropping behavior diminishes foraging efficiency of a predatory ladybeetle**. *Entomologia Experimentalis Et Applicata* 2008, **127**: 118-124.
7. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM: **Variable host-pathogen compatibility in *Mycobacterium tuberculosis***. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**: 2869-2873.

8. Govind S: **Innate immunity in *Drosophila*: Pathogens and pathways**. *Insect Science* 2008, **15**: 29-43.

9. Strand MR: **The insect cellular immune response**. *Insect Science* 2008, **15**: 1-14.

10. Schulenburg H, Boehnisch C, Michiels NK: **How do invertebrates generate a highly specific innate immune response?** *Molecular Immunology* 2007, **44**: 3338-3344.

11. Boutros M, Agaisse H, Perrimon N: **Sequential activation of signaling pathways during innate immune responses in *Drosophila***. *Developmental Cell* 2002, **3**: 711-722.

12. Dionne MS, Schneider DS: **Models of infectious diseases in the fruit fly *Drosophila melanogaster***. *Disease Models & Mechanisms* 2008, **1**: 43-49.

13. Zou Z, Evans JD, Lu ZQ, Zhao PC, Williams M, Sumathipala N, Hetru C, Hultmark D, Jiang HB: **Comparative genomic analysis of the *Tribolium* immune system**. *Genome Biology* 2007, **8**.

14. Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, Kanost M, Thompson GJ, Zou Z, Hultmark D: **Immune pathways and defence mechanisms in honey bees *Apis mellifera***. *Insect Molecular Biology* 2006, **15**: 645-656.

15. Waterhouse RM, Kriventseva EV, Meister S, Xi ZY, Alvarez KS, Bartholomay LC, Barillas-Mury C, Bian GW, Blandin S, Christensen BM, Dong YM, Jiang HB, Kanost MR, Koutsos AC, Levashina EA, Li JY, Ligoxygakis P, MacCallum RM, Mayhew GF, Mendes A, Michel K, Osta MA, Paskewitz S, Shin SW, Vlachou D, Wang LH, Wei WQ, Zheng LB, Zou

Z, Severson DW *et al*: **Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes**. *Science* 2007, **316**: 1738-1743.

16. Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G, Hetru C, Hoa N, Hoffmann JA, Kanzok SM, Letunic I, Levashina EA, Loukeris TG, Lycett G, Meister S, Michel K, Muller HM, Osta MA, Paskewitz SM, Reichhart JM, Rzhetsky A, Troxler L, Vernick KD, Vlachou D, Volz J, von Mering C *et al*: **Immunity-related genes and gene families in *Anopheles gambiae***. *Science* 2002, **298**: 159-165.

17. Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG: **Dynamic evolution of the innate immune system in *Drosophila***. *Nature Genetics* 2007, **39**: 1461-1468.

18. Pockley AG: **Heat shock proteins as regulators of the immune response**. *Lancet* 2003, **362**: 469-476.

19. Kunert G, Otto S, Rose USR, Gershenzon J, Weisser WW: **Alarm pheromone mediates production of winged dispersal morphs in aphids**. *Ecology Letters* 2005, **8**: 596-603.

20. Moran NA, Russell JA, Koga R, Fukatsu T: **Evolutionary relationships of three new species of Enterobacteriaceae living as symbionts of aphids and other insects**. *Applied and Environmental Microbiology* 2005, **71**: 3302-3310.

21. Moran NA, Telang A: **Bacteriocyte-associated symbionts of insects - a variety of insect groups harbor ancient prokaryotic endosymbionts**. *Bioscience* 1998, **48**: 295-304.

22. Sandstrom JP, Russell JA, White JP, Moran NA: **Independent origins and horizontal transfer of bacterial symbionts of aphids**. *Molecular Ecology* 2001, **10**: 217-228.

23. Tsuchida T, Koga R, Fukatsu T: **Host plant specialization governed by facultative symbiont**. *Science* 2004, **303**: 1989-1989.

24. Russell JA, Moran NA: **Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures**. *Proceedings of the Royal Society B-Biological Sciences* 2006, **273**: 603-610.

25. Montllor CB, Maxmen A, Purcell AH: **Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress**. *Ecological Entomology* 2002, **27**: 189-195.

26. Chen DQ, Montllor CB, Purcell AH: **Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid, *A. kondoi***. *Entomologia Experimentalis Et Applicata* 2000, **95**: 315-323.

27. Oliver KM, Moran NA, Hunter MS: **Variation in resistance to parasitism in aphids is due to symbionts not host genotype**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**: 12795-12800.

28. Oliver KM, Russell JA, Moran NA, Hunter MS: **Facultative bacterial symbionts in aphids confer resistance to parasitic wasps**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**: 1803-1807.

29. Scarborough CL, Ferrari J, Godfray HCJ: **Aphid protected from pathogen by endosymbiont**. *Science* 2005, **310**: 1781-1781.

30. Haine ER: **Symbiont-mediated protection**. *Proceedings of the Royal Society B-Biological Sciences* 2008, **275**: 353-361.

31. Altincicek B, Gross J, Vilcinskas A: **Wounding-mediated gene expression and accelerated viviparous reproduction of the pea aphid *Acyrthosiphon pisum***. *Insect Molecular Biology* 2008, **17**: 711-716.

32. Altincicek B, Knorr E, Vilcinskas A: **Beetle immunity: Identification of immune-inducible genes from the model insect *Tribolium castaneum***. *Developmental and Comparative Immunology* 2008, **32**: 585-595.

33. Altincicek B, Vilcinskas A: **Analysis of the immune-inducible transcriptome from microbial stress resistant, rat-tailed maggots of the drone fly *Eristalis tenax***. *Bmc Genomics* 2007, **8**.

34. Altincicek B, Vilcinskas A: **Identification of immune-related genes from an apterygote insect, the firebrat *Thermobia domestica***. *Insect Biochemistry and Molecular Biology* 2007, **37**: 726-731.

35. Anselme C, Perez-Brocal V, Vallier A, Vincent-Monegat C, Charif D, Latorre A, Moya A, Heddi A: **Identification of the Weevil immune genes and their expression in the bacteriome tissue**. *Bmc Biology* 2008, **6**.

36. Consortium IAG: **Genome sequence of the pea aphid, *Acyrthosiphon pisum***. *Submitted to PLOS Biology* 2009.

37. Kaneko T, Silverman N: **Bacterial recognition and signalling by the *Drosophila* IMD pathway**. *Cellular Microbiology* 2005, **7**: 461-469.

38. Steiner H: **Peptidoglycan recognition proteins: on and off switches for innate immunity**. *Immunological Reviews* 2004, **198**: 83-96.

39. Werner T, Liu G, Kang D, Ekengren S, Steiner H, Hultmark D: **A family of peptidoglycan recognition proteins in the fruit fly *Drosophila***

melanogaster. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**: 13772-13777.

40. McTaggert SJ, Conlon C, Colbourne JK, Blaxter ML, Little TJ: **The components of the *Daphnia pulex* immune system as revealed by complete genome sequencing**. *BMC Genomics* 2009, **10**: 175.

41. Lemaitre B, Hoffmann J: **The host defense of *Drosophila melanogaster***. *Annual Review of Immunology* 2007, **25**: 697-743.

42. Goftar M, Gobert V, Matskevich AA, Reichhart JM, Wang CS, Buft TM, BeIvin M, Hoffmann JA, Ferrandon D: **Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors**. *Cell* 2006, **127**: 1425-1437.

43. Tanji T, Ohashi-Kobayashi A, Natori S: **Participation of a galactose-specific C-type lectin in *Drosophila* immunity**. *Biochemical Journal* 2006, **396**: 127-138.

44. Ao JQ, Ling EJ, Yu XQ: ***Drosophila* C-type lectins enhance cellular encapsulation**. *Molecular Immunology* 2007, **44**: 2541-2548.

45. Pace KE, Baum LG: **Insect galectins: Roles in immunity and development**. *Glycoconjugate Journal* 2004, **19**: 607-614.

46. Dimopoulos G, Seeley D, Wolf A, Kafatos FC: **Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle**. *Embo Journal* 1998, **17**: 6115-6123.

47. Dimopoulos G, Richman A, dellaTorre A, Kafatos FC, Louis C: **Identification and characterization of differentially expressed cDNAs of**

the vector mosquito, *Anopheles gambiae*. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**: 13066-13071.

48. Ramet M, Pearson A, Manfruelli P, Li XH, Koziel H, Gobel V, Chung E, Krieger M, Ezekowitz RAB: ***Drosophila* scavenger receptor Cl is a pattern recognition receptor for bacteria**. *Immunity* 2001, **15**: 1027-1038.

49. Lazzaro BP: **Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans***. *Genetics* 2005, **169**: 2023-2034.

50. Ju JS, Cho MH, Brade L, Kim JH, Park JW, Ha NC, Soderhall I, Soderhall K, Brade H, Lee BL: **A novel 40-kDa protein containing six repeats of an epidermal growth factor-like domain functions as a pattern recognition protein for lipopolysaccharide**. *Journal of Immunology* 2006, **177**: 1838-1845.

51. Kurucz E, Markus R, Zsamboki J, Folkl-Medzihradszky K, Darula Z, Vilmos P, Udvardy A, Krausz I, Lukacsovich T, Gateff E, Zettervall CJ, Hultmark D, Ando I: **Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes**. *Current Biology* 2007, **17**: 649-654.

52. Somogyi K, Sipos B, Penzes Z, Kurucz E, Zsamboki J, Hultmark D, Ando I: **Evolution of Genes and Repeats in the Nimrod Superfamily**. *Molecular Biology and Evolution* 2008, **25**: 2337-2347.

53. Dong YM, Taylor HE, Dimopoulos G: **AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system**. *Plos Biology* 2006, **4**: 1137-1146.

54. Brites D, McTaggart S, Morris K, Anderson J, Thomas K, Colson I, Fabbro T, Little TJ, Ebert D, Du Pasquier L: **The Dscam homologue of the crustacean**

***Daphnia* is diversified by alternative splicing like in insects**. *Molecular Biology and Evolution* 2008, **25**: 1429-1439.

55. Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, Rebel VI, Schmucker D: **Extensive diversity of Ig-superfamily proteins in the immune system of insects**. *Science* 2005, **309**: 1874-1878.

56. Schlenke TA, Morales J, Govind S, Clark AG: **Contrasting infection strategies in generalist and specialist wasp parasitoids of *Drosophila melanogaster***. *PloS Pathogens* 2007, **3**: 1486-1501.

57. Kanost MR: **Serine proteinase inhibitors in arthropod immunity**. *Developmental and Comparative Immunology* 1999, **23**: 291-301.

58. Krem MM, Di Cera E: **Evolution of enzyme cascades from embryonic development to blood coagulation**. *Trends in Biochemical Sciences* 2002, **27**: 67-74.

59. Rawlings ND, Barrett AJ: **Evolutionary families of peptidases**. *Biochemical Journal* 1993, **290**: 205-218.

60. Imler JL, Hoffmann JA: **Toll receptors in *Drosophila*: a family of molecules regulating development and immunity**. In: *Toll-Like Receptor Family Members and Their Ligands.* vol. 270; 2002: 63-79.

61. Leulier F, Lemaitre B: **Toll-like receptors - taking an evolutionary approach**. *Nature Reviews Genetics* 2008, **9**: 165-178.

62. Tauszig S, Jouanguy E, Hoffmann JA, Imler JL: **Toll-related receptors and the control of antimicrobial peptide expression in *Drosophila***. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**: 10520-10525.

63. Agaisse H, Perrimon N: **The roles of JAK/STAT signaling in *Drosophila* immune responses**. *Immunological Reviews* 2004, **198**: 72-82.

64. Lemaitre B, Kromermetzger E, Michaut L, Nicolas E, Meister M, Georgel P, Reichhart JM, Hoffmann JA: **A recessive mutation, immume-deficiency (IMD), defines 2 distinct control pathways in the *Drosophila* host-defense**. *Proceedings of the National Academy of Sciences of the United States of America* 1995, **92**: 9465-9469.

65. Delaney JR, Stoven S, Uvell H, Anderson KV, Engstrom Y, Mlodzik M: **Cooperative control of *Drosophila* immune responses by the JNK and NF-kappa B signaling pathways**. *Embo Journal* 2006, **25**: 3068-3077.

66. Bidla G, Dushay MS, Theopold U: **Crystal cell rupture after injury in *Drosophila* requires the JNK pathway, small GTPases and the TNF homolog Eiger**. *Journal of Cell Science* 2007, **120**: 1209-1215.

67. Igaki T, Kanda H, Yamamoto-Goto Y, Kanuka H, Kuranaga E, Aigaki T, Miura M: **Eiger, a TNF superfamily ligand that triggers the *Drosophila* JNK pathway**. *Embo Journal* 2002, **21**: 3009-3018.

68. Bulet P, Stocklin R: **Insect antimicrobial peptides: Structures, properties and gene regulation**. *Protein and Peptide Letters* 2005, **12**: 3-11.

69. Bulet P, Stocklin R, Menin L: **Anti-microbial peptides: from invertebrates to vertebrates**. *Immunological Reviews* 2004, **198**: 169-184.

70. Lemaitre B, Reichhart JM, Hoffmann JA: ***Drosophila* host defense: Differential induction of antimicrobial peptide genes after infection by various classes of microorganisms**. *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**: 14614-14619.

71. Hoffmann JA, Reichhart JM: ***Drosophila*** **innate immunity: an evolutionary perspective**. *Nature Immunology* 2002, **3**: 121-126.

72. Shatters RG, Boykin LM, Lapointe SL, Hunter WB, Weathersbee AA: **Phylogenetic and structural relationships of the PR5 gene family reveal an ancient multigene family conserved in plants and select animal taxa**. *Journal of Molecular Evolution* 2006, **63**: 12-29.

73. Jolles P: **Lysozymes: Model Enzymes in Biochemistry and Biology**. Basel, Switzerland: Birkhauser; 1996.

74. Regel R, Matioli SR, Terra WR: **Molecular adaptation of *Drosophila melanogaster* lysozymes to a digestive function**. *Insect Biochemistry and Molecular Biology* 1998, **28**: 309-319.

75. Li B, Calvo E, Marinotti O, James AA, Paskewitz SM: **Characterization of the c-type lysozyme gene family in *Anopheles gambiae***. *Gene* 2005, **360**: 131-139.

76. Paskewitz SM, Li B, Kajla MK: **Cloning and molecular characterization of two invertebrate-type lysozymes from *Anopheles gambiae***. *Insect Molecular Biology* 2008, **17**: 217-225.

77. Nakabachi A, Shigenobu S, Sakazume N, Shiraki T, Hayashizaki Y, Carninci P, Ishikawa H, Kudo T, Fukatsu T: **Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera***. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**: 5477-5482.

78. Daimon T, Hamada K, Mita K, Okano K, Suzuki MG, Kobayashi M, Shimada T: **A *Bombyx mori* gene, BmChi-h, encodes a protein homologous to**

bacterial and baculovirus chitinases. *Insect Biochem Mol Biol* 2003, **33**: 749-759.

79. Nakabachi A, Shigenobu S, Miyagishima S: **Chitinase-like proteins encoded in the genome of the pea aphid, *Acyrthosiphon pisum***. *Insect Molecular Biology* 2009, **in press**.

80. Lagueux M, Perrodou E, Levashina EA, Capovilla M, Hoffmann JA: **Constitutive expression of a complement-like protein in Toll and JAK gain-of-function mutants of *Drosophila***. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**: 11427-11432.

81. Agaisse H, Petersen UM, Boutros M, Mathey-Prevot B, Perrimon N: **Signaling role of hemocytes in *Drosophila* JAK/STAT-dependent response to septic injury**. *Developmental Cell* 2003, **5**: 441-450.

82. Ekengren S, Hultmark D: **A family of Turandot-related genes in the humoral stress response of *Drosophila***. *Biochemical and Biophysical Research Communications* 2001, **284**: 998-1003.

83. Ekengren S, Tryselius Y, Dushay MS, Liu G, Steiner H, Hultmark D: **A humoral stress response in *Drosophila***. *Current Biology* 2001, **11**: 714-718.

84. Nappi AJ, Christensen BM: **Melanogenesis and associated cytotoxic reactions: Applications to insect innate immunity**. *Insect Biochemistry and Molecular Biology* 2005, **35**: 443-459.

85. Soderhall K, Cerenius L: **Role of the prophenoloxidase-activating system in invertebrate immunity**. *Current Opinion in Immunology* 1998, **10**: 23-28.

86. Foley E, O'Farrell PH: **Nitric oxide contributes to induction of innate immune responses to Gram-negative bacteria in *Drosophila***. *Genes & Development* 2003, **17**: 115-125.

87. Rivero A: **Nitric oxide: an antiparasitic molecule of invertebrates**. *Trends in Parasitology* 2006, **22**: 219-225.

88. Henderson B, Allan E, CoateS ARM: **Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection**. *Infection and Immunity* 2006, **74**: 3693-3706.

89. Aguilar R, Jedlicka AE, Mintz M, Mahairaki V, Scott AL, Dimopoulos G: **Global gene expression analysis of *Anopheles gambiae* responses to microbial challenge**. *Insect Biochemistry and Molecular Biology* 2005, **35**: 709-719.

90. Dimopoulos G, Christophides GK, Meister S, Schultz J, White KP, Barillas-Mury C, Kafatos FC: **Genome expression analysis of *Anopheles gambiae*: Responses to injury, bacterial challenge, and malaria infection**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**: 8814-8819.

91. Guedes SD, Vitorino R, Domingues R, Tomer K, Correia AJF, Amado F, Domingues P: **Proteomics of immune-challenged *Drosophila melanogaster* larvae hemolymph**. *Biochemical and Biophysical Research Communications* 2005, **328**: 106-115.

92. Ursic-Bedoya RJ, Lowenberger CA: ***Rhodnius prolixus*: Identification of immune-related genes up-regulated in response to pathogens and parasites using suppressive subtractive hybridization**. *Developmental and Comparative Immunology* 2007, **31**: 109-120.

93. Girardot F, Monnier V, Tricoire H: **Genome wide analysis of common and specific stress responses in adult drosophila melanogaster**. *Bmc Genomics* 2004, **5**.

94. Rudenko N, Golovchenko M, Edwards MJ, Grubhoffer L: **Differential expression of *Ixodes ricinus* tick genes induced by blood feeding or *Borrelia burgdorferi* infection**. *Journal of Medical Entomology* 2005, **42**: 36-41.

95. Ramsey JS, Rider DS, Walsh T, de Vos M, Gordon K, Ponnala L, Roe BA, Jander G: **Comparative analysis of detoxification enzymes in *Acrythrosiphon pisum* and *Myzus persicae***. *Insect Molecular Biology* 2009, **in press**.

96. Kislow CJ, Edwards LJ: **Repellent odor in aphids**. *Nature* 1972, **235**: 108-109.

97. Xiangyu JG, Zhang F, Fang YL, Kan W, Zhang GX, Zhang ZN: **Behavioural response of aphids to the alarm pheromone component (E)-beta-farnesene in the field**. *Physiological Entomology* 2002, **27**: 307-311.

98. Lewis MJ, Prosser IM, Mohib A, Field LM: **Cloning and characterisation of a prenyltransferase from the aphid Myzus persicae with potential involvement in alarm pheromone biosynthesis**. *Insect Molecular Biology* 2008, **17**: 437-443.

99. Grenier AM, Duport G, Pages S, Condemine G, Rahbe Y: **The phytopathogen *Dickeya dadantii* (*Erwinia chrysanthemi* 3937) is a pathogen of the pea aphid**. *Applied and Environmental Microbiology* 2006, **72**: 1956-1965.

100. Gaunt MW, Miles MA: **An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks**. *Molecular Biology and Evolution* 2002, **19**: 748-761.

101. Bensadia F, Boudreault S, Guay JF, Michaud D, Cloutier C: **Aphid clonal resistance to a parasitoid fails under heat stress**. *Journal of Insect Physiology* 2006, **52**: 146-157.

102. Carver M, Sullivan DJ: **Encapsulative defence reactions of aphids (Hemiptera: Aphididae) to insect parasitoids (Hymenoptera: Aphidiidae and Aphelinidae)**. In: *Ecology and Effectiveness of Aphidophaga.* Edited by Niemczyk E, Dixon AFG. The Hague: SPB Academic Publishing; 1988.

103. Douglas AE: **Phloem-sap feeding by animals: problems and solutions**. In: *Annual Meeting of the Society-for-Experimental-Biology: Jul 2005; Barcelona, Spain*; 2005: 747-754.

104. Stavrinides J, McCloskey JK, Ochman H: **The pea aphid as both host and vector for the phytopathogenic bacterium, *Pseudomonas syringae***. *Applied Environmental Microbiology* 2009, **75**: 2230-2235.

105. Nakabachi A, Ishikawa H, Kudo T: **Extraordinary proliferation of microorganisms in aposymbiotic pea aphids, *Acyrthosiphon pisum***. *Journal of Invertebrate Pathology* 2003, **82**: 152-161.

106. Minchella DJ, Loverde PT: **A cost of increased early reproductive effort in the snail *Biomphalaria glabarata***. *American Naturalist* 1981, **118**: 876-881.

107. Minchella DJ, Leathers BK, Brown KM, McNair JN: **Host and parasite counteradaptations - an example from a fresh-water snail**. *American Naturalist* 1985, **126**: 843-854.

108. Adamo SA: **Evidence for adaptive changes in egg laying in crickets exposed to bacteria and parasites**. *Animal Behaviour* 1999, **57**: 117-124.

109. Chadwick W, Little TJ: **A parasite-mediated life-history shift in *Daphnia magna***. *Proceedings of the Royal Society B-Biological Sciences* 2005, **272**: 505-509.

110. Polak M, Starmer WT: **Parasite-induced risk of mortality elevates reproductive effort in male *Drosophila***. *Proceedings of the Royal Society of London Series B-Biological Sciences* 1998, **265**: 2197-2201.

111. Barribeau SM, Sok D, Linares C, Gerardo NM: **Bacterial symbionts alter aphid reproductive investment in response to mortality risks**. *Evolution* submitted.

112. Baverstock J, Roy HE, Clark SJ, Alderson PG, Pell JK: **Effect of fungal infection on the reproductive potential of aphids and their progeny**. *Journal of Invertebrate Pathology* 2006, **91**: 136-139.

113. MacArthur RH, Wilson EO: **The Theory of Island Biogeography**. Princeton, N.J.: Princeton University Press; 1967.

114. Zuk M, Stoehr AM: **Immune defense and host life history**. *American Naturalist* 2002, **160**: S9-S22.

115. Miller MR, White A, Boots M: **Host life span and the evolution of resistance characteristics**. *Evolution* 2007, **61**: 2-14.

116. Pianka ER: **On r- and K-selection**. *American Naturalist* 1970, **104**: 592-597.

117. Fukatsu T, Nikoh N, Kawai R, Koga R: **The secondary endosymbiotic bacterium of the pea aphid *Acyrthosiphon pisum* (Insecta : Homoptera)**. *Applied and Environmental Microbiology* 2000, **66**: 2748-2758.

118. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic Local Alignment Search Tool**. *Journal of Molecular Biology* 1990, **215**: 403-410.

119. Thompson JD, Higgins DG, Gibson TJ: **Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Research* 1994, **22**: 4673-4680.

120. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput**. *Nucleic Acids Research* 2004, **32**: 1792-1797.

121. Eddy SR: **Profile hidden Markov models**. *Bioinformatics* 1998, **14**: 755-763.

122. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer ELL: **The Pfam Protein Families Database**. *Nucleic Acids Research* 2002, **30**: 276-280.

123. Huerta-Cepas J, Bueno A, Dopazo JQ, Gabaldon T: **PhylomeDB: a database for genome-wide collections of gene phylogenies**. *Nucleic Acids Research* 2008, **36**: D491-D496.

124. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood**. *Systematic Biology* 2003, **52**: 696-704.

125. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R: **Dendroscope: An interactive viewer for large phylogenetic trees**. *Bmc Bioinformatics* 2007, **8**.

126. Sabater-Munoz B, Legeai F, Rispe C, Bonhomme J, Dearden P, Dossat C, Duclert A, Gauthier JP, Ducray DG, Hunter W, Dang P, Kambhampati S, Martinez-Torres D, Cortes T, Moya A, Nakabachi A, Philippe C, Prunier-Leterme N, Rahbe Y, Simon JC, Stern DL, Wincker P, Tagu D: **Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera)**. *Genome Biology* 2006, **7**.

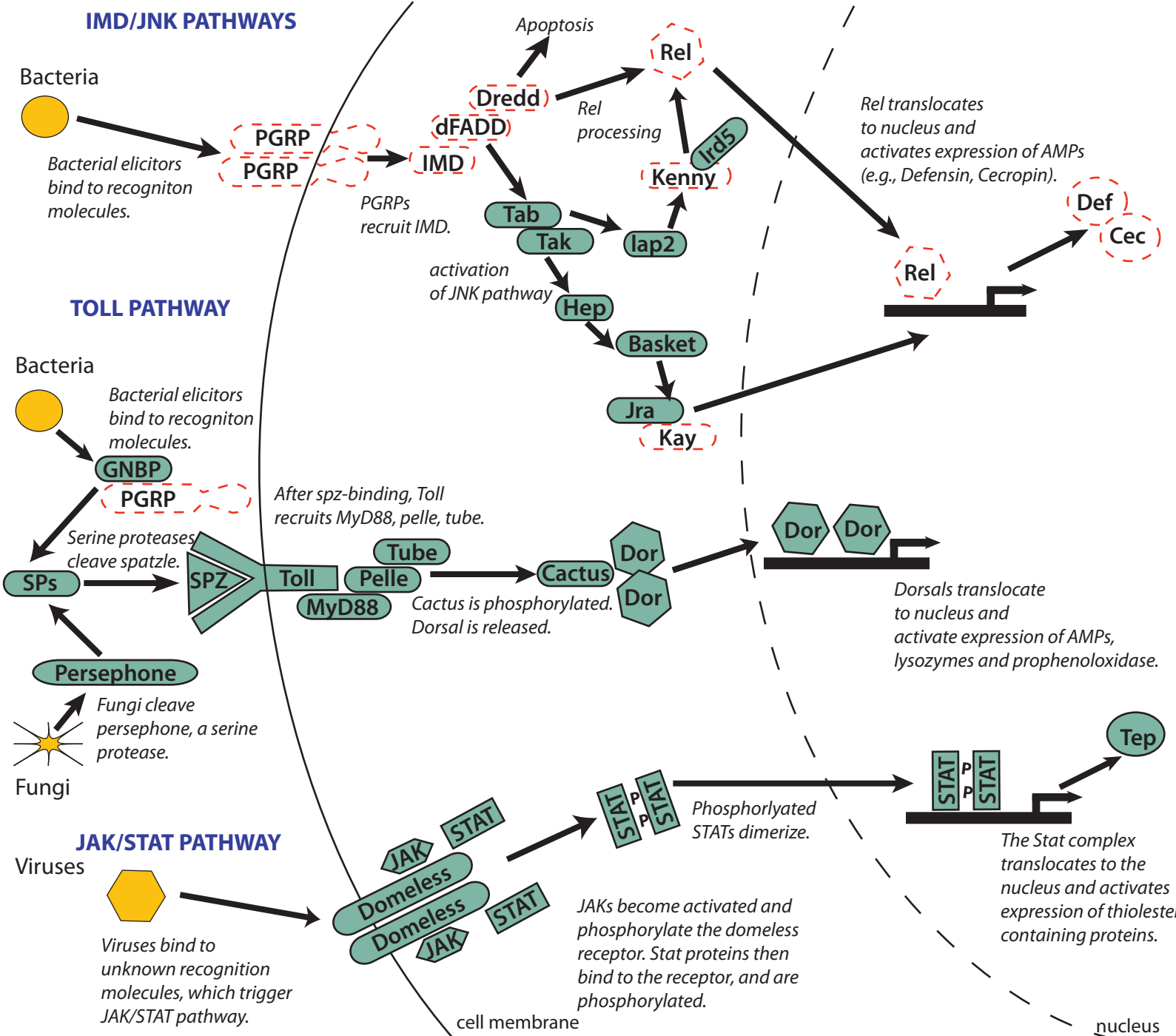
127. Rahbe Y, Delobel B, Febvay G, Chantegrel B: **Aphid-specific triglycerides in symbiotic and aposymbiotic *Acyrthosiphon pisum***. *Insect Biochemistry and Molecular Biology* 1994, **24**: 95-101.

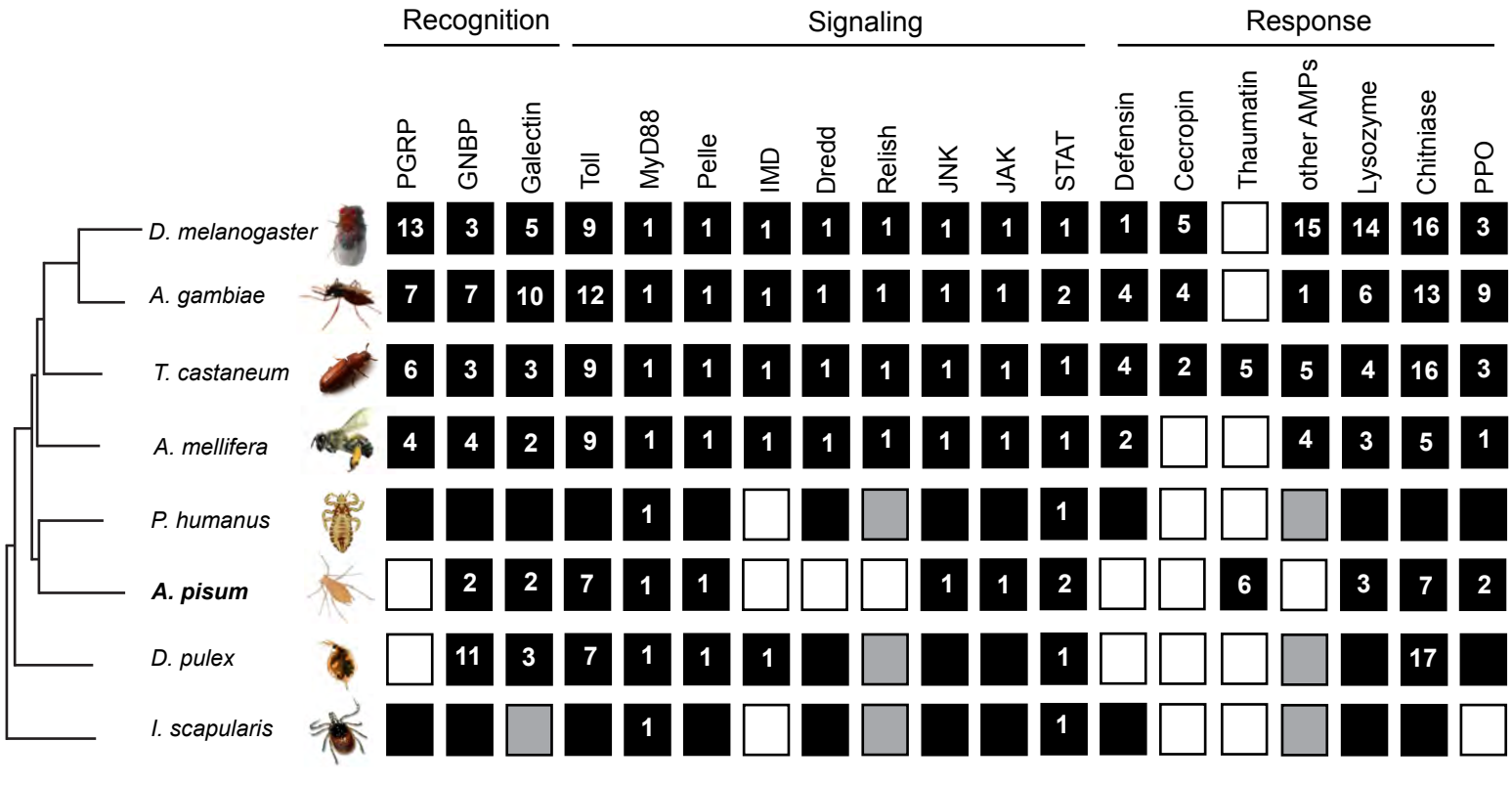
128. Chernysh S, Cociancich S, Briand JP, Hetru C, Bulet P: **The inducible antibacterial peptides of the hemipteran insect *Palomena prasina*: Identification of a unique family of proline-rich peptides and of a novel insect defensin**. *Journal of Insect Physiology* 1996, **42**: 81-89.

129. Sonenshine DE, Hynes WL: **Molecular characterization and related aspects of the innate immune response in ticks**. *Frontiers in Bioscience* 2008, **13**: 7046-7063.

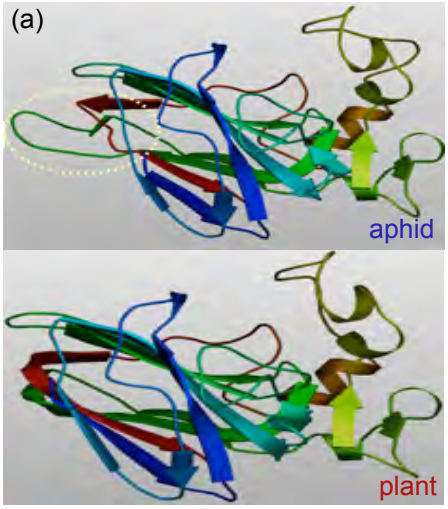
130. Dall'Antonia Y, Pavkov T, Fuchs H, Breiteneder H, Keller W: **Crystallization and preliminary structure determination of the plant food allergen Pru av 2**. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications* 2005, **61**: 186-188.

131. **Swissmodel** [http://swissmodel.expasy.org/workspace]


IMD/JNK PATHWAYS
Bacteria
Bacterial elicitors bind to recogniton molecules.
PGRP
PGRP
IMD
dFADD
Dredd
Apoptosis
Rel
Rel processing
Kenny
Ird5
PGRPs recruit IMD.
Tab
Tak
Iap2
activation of JNK pathway
Hep
Basket
Jra
Kay
Rel translocates to nucleus and activates expression of AMPs (e.g., Defensin, Cecropin).
Def
Cec
Rel
TOLL PATHWAY
Bacteria
Bacterial elicitors bind to recogniton molecules.
GNBP
PGRP
After spz-binding, Toll recruits MyD88, pelle, tube.
Serine proteases cleave spatzle.
SPs
SPZ
Toll
Tube
Pelle
MyD88
Cactus
Dor
Dor
Cactus is phosphorylated. Dorsal is released.
Dor
Dor
Dorsals translocate to nucleus and activate expression of AMPs, lysozymes and prophenoloxidase.
Persephone
Fungi cleave persephone, a serine protease.
Fungi
JAK/STAT PATHWAY
Viruses
Viruses bind to unknown recognition molecules, which trigger JAK/STAT pathway.
JAK
STAT
Domeless
Domeless
STAT
JAK
JAKs become activated and phosphorylate the domeless receptor. Stat proteins then bind to the receptor, and are phosphorylated.
STAT
P
P
STAT
Phosphorlyated STATs dimerize.
STAT
P
P
STAT
Tep
The Stat complex translocates to the nucleus and activates expression of thiolester containing proteins.
cell membrane
nucleus


Recognition
Signaling
Response
PGRP
GNBP
Galectin
Toll
MyD88
Pelle
IMD
Dredd
Relish
JNK
JAK
STAT
Defensin
Cecropin
Thaumatin
other AMPs
Lysozyme
Chitniase
PPO
D. melanogaster 13 3 5 9 1 1 1 1 1 1 1 1 1 5 15 14 16 3
A. gambiae 7 7 10 12 1 1 1 1 1 1 1 2 4 4 1 6 13 9
T. castaneum 6 3 3 9 1 1 1 1 1 1 1 1 4 2 5 5 4 16 3
A. mellifera 4 4 2 9 1 1 1 1 1 1 1 1 2 4 3 5 1
P. humanus 1
A. pisum 2 2 7 1 1 1 1 2 6 3 7 2
D. pulex 11 3 7 1 1 1 1 17
I. scapularis 1

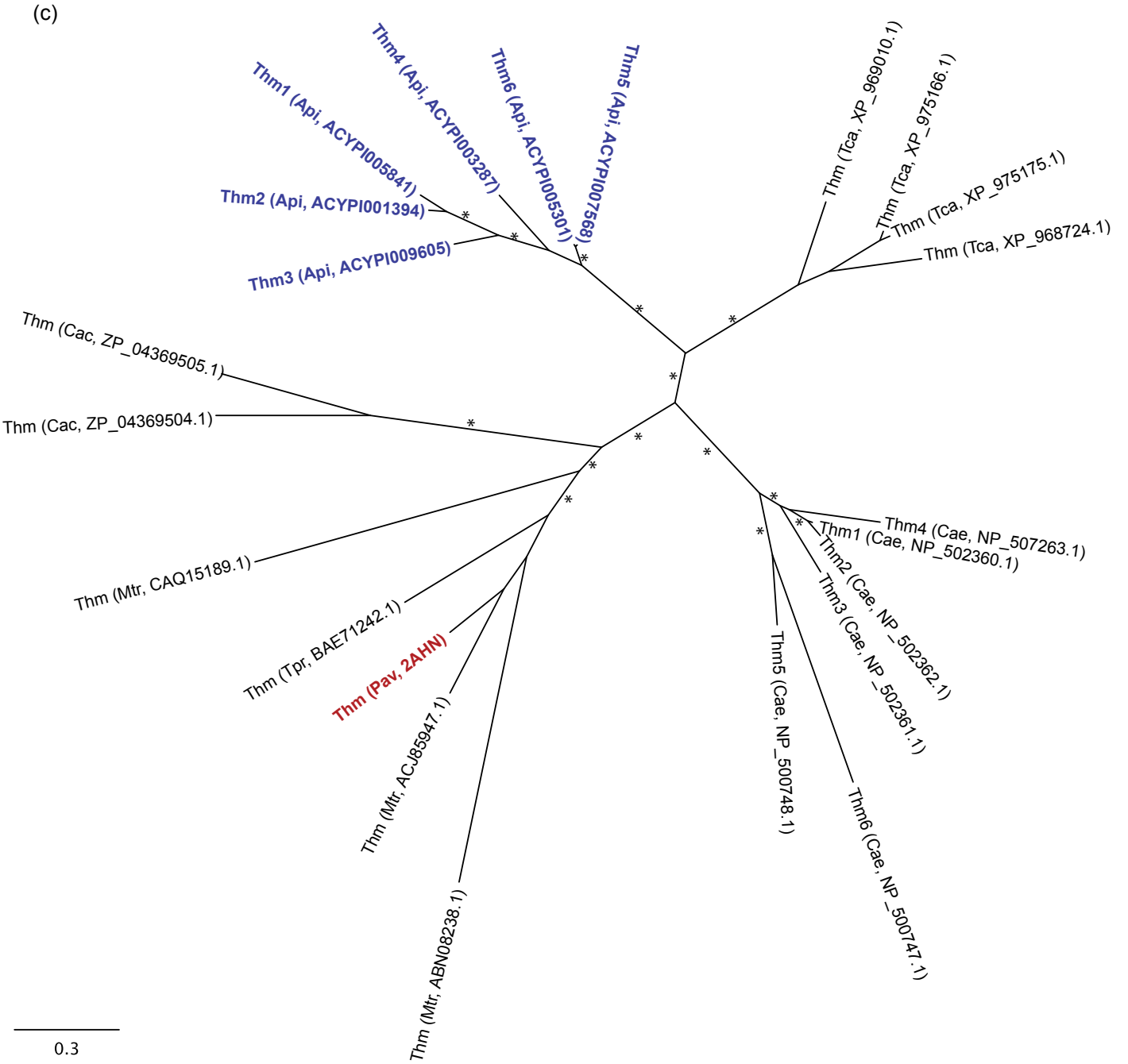
(a)

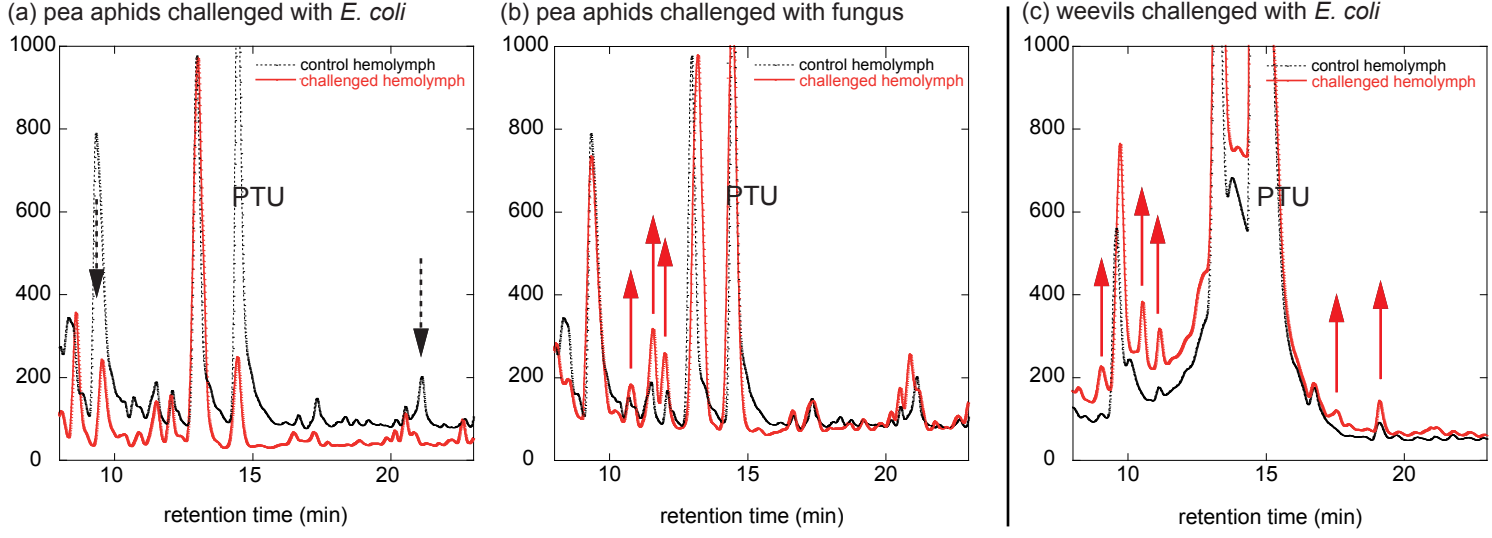


(b)

```
PLANT                                ATISFKNNCPYMVWPGTLTSDQKPQLSTTGFEL
APHID   MDFKFILFFVQMSGVTMAHMIRVTNNCPFTVWPGIQGNTGQQHLENGGFSV

PLANT   ASQASFQLDTPVPWNGRFWARTGCSTDASGKFVCATADCASGQVMCNGNGA
APHID   GAYKTHFILSSRNWAGRIWGRTNC--DSQGK--CETGDCGN-KIQCNGTLG

PLANT   IPPATLAEFNIPAGGGQDFYDVSLVDGFNLPMSVTPQG------GTGDCKT
APHID   VPPLTLAEIQFAESDNIDSYYVSLVDGFNLPIKIMPNKYPMTSKNSIDCKP

PLANT   ASCPANVNAVCPSELQKKGSDGS-VVACLSACVKFGTPQYCCTPPQNTPET
APHID   ADCVADLNSKCPDKLAVKAADGSSVVACKSACALFNTDSDCCQGVYTTPAT

PLANT   CP----PTNYSEIFHNACPDAYSYAYDDKRGTFTCNGGP--NYAITFCP
APHID   CNSSSWPQNYPPFFKKACPYAYSYPFDNTTSTFTCHGNSLTKFDIVFCP
```

(c)




(a) pea aphids challenged with E. coli
(b) pea aphids challenged with fungus
(c) weevils challenged with E. coli
control hemolymph
challenged hemolymph
PTU
retention time (min)
0
200
400
600
800
1000
10
15
20

Supplemental Material: Immunity and defense in pea aphids, *Acyrthosiphon pisum*

Supplementary Methods for Gene Expression Study

Gene expression overview. We utilized real-time quantitative PCR to conduct a preliminary investigation of the expression of 23 recognition, signaling and response genes in aphids subjected to a number of infection and stress treatments (Tables S2, below). First, we verified expression of the immune and stress genes in aphids stabbed with needles inoculated with the natural commensal *Escherichia coli*, a natural Gram-positive bacterial pathogen, a natural Gram-negative bacterial pathogen, or with no needle (control). More closely mimicking the uptake of bacteria while feeding, we verified expression of the genes in aphids fed on artificial diet containing either *E. coli* or a Gram-negative pathogen relative to that of aphids fed on control diet with no bacteria. We also verified gene expression in entomopathogen fungus-infected to uninfected aphids, and in aphids stabbed with pathogenic virus-inoculated needles to unstabbed control aphids. All microbes used have been isolated from pea aphids in natural populations. Finally, to assess effects of environmental stress, we verified gene expression in aphids exposed to alarm pheromones, starvation, and heat-shock.

We conducted the five challenge experiments with aphids of clonal line LSR1, the clone sequenced in the genome sequencing project. This line was cured of its secondary symbiont, *R. insecticola*, more than two years ago using a standard antibiotic treatment [1], and has been continuously maintained as an asexual clone on fava beans at 16hr light: 8hr dark at 20^{o}C. For each experiment, we used 5-6 day old, unwinged aphids, which were maintained on fava bean plants at 20^{o}C unless otherwise noted.

Bacterial stabbing experiment. For the first experiment, we stabbed aphids with *E. coli*, Gram-negative pathogen Ng5B, in the genus *Enterobacter*, or Gram-positive pathogen 6B, in the genus *Staphylococcus*. *E. coli* has been previously shown to be a commensal that has little affect on aphid

survival when either injected into or fed to aphids as it is quickly cleared by the host (Figure S3, below) [2]. Ng5b and 6b are bacteria originally isolated from laboratory pea aphids. When fed to aphids, these pathogens kill most aphids in approximately 48 hours (Figure S3a, below). The night before the infection, we grew bacterial cultures from glycerol stocks on Luria broth (LB) agar at 37^{o}C. The morning of the infection, we transferred bacterial colonies to LB and grew them at 37^{o}C. We determined concentration of broth cultures by optical density (OD_{600}), and then standardized the cultures to OD_{600} = 0.5. Next, we stabbed aphids with a minutin pin contaminated with the standardized bacterial cultures or with sterile LB and then transferred the stabbed aphids and unstabbed control aphids to fava bean plants. After eight hours, we froze five aphids per condition in liquid nitrogen and stored them at -80^{o}C for subsequent RNA extraction. We monitored ten aphids per condition for survival (Figure S3, below).

Bacterial feeding experiment. For the second experiment, we fed aphids on AP3 artificial *A. pisum* diet [3] containing no bacteria (control), *E. coli*, or the Gram-negative pathogen Ng5b. We grew bacteria as above and then inoculated the treatment diets with 1 μL bacterial culture per 20 mL diet, and control diet with the same amount of bacteria-free LB. We plated the diets onto LB agar to confirm that the control diet was free of bacteria and that the final bacterial concentration in the treatment diet was approximately 2 x 10^4 colony forming units (cfu) per mL. We added one drop of food-grade blue dye per 5 mL of diet to allow us to detect whether aphids had ingested the diet. We then filled 10 mm Petri dish bottoms with diet (either control or treatment) and covered the dish with stretched Parafilm. We affixed the feeding dishes to the bottom of 15 mm Petri dishes, and transferred 30 - 50 aphids directly from plants to each feeding dish. We maintained dishes upside down in a 20^{o}C incubator. Approximately 12 hours after being exposed to the diet, we froze five aphids from each treatment that had fed (as determined by the presence of dye in the aphid intestinal tract) in liquid nitrogen and maintained them at -80^{o}C for subsequent RNA extraction. Eight hours later, we transferred the

remaining aphids that had fed (n = 30 per treatment) to fava bean plants and monitored them for survival (Figure S3a, below).

Fungal shower experiment. For the third experiment, we exposed aphids to a shower of spores of the fungus *Zoophthora occidentalis,* an aphid specific fungal entomopathogen [4]. We placed aphids in a 70 mm tall cylinder with a damp sponge at the bottom. We then inverted an approximately two-week-old culture of *Z. occidentalis* on potato dextrose agar (PDA) over the cylinder and allowed the spores to fall on the aphids for 2 hours. Control aphids were exposed to the same conditions but with a sterile PDA plate. After exposure, we transferred the aphids to fava plants. We froze five aphids per condition in liquid nitrogen 24 hours after exposure and monitored 22 aphids per condition for survival (Figure S3b, below). The frozen samples were maintained at -80°C for subsequent RNA extraction.

Viral stabbing experiment. For the fourth experiment, we stabbed aphids with a solution containing Aphid Lethal Paralysis Virus (ALPV), a RNA virus found in natural populations that is lethal to pea aphids (Georgievska, Miller & Bonning, unpublished data) [5, 6]. The solution was made by grinding up virus-killed aphid cadavers in 7.5 μL of water per cadaver. After stabbing, aphids were isolated for an hour and then kept on fava bean plants until 16 hours from initial stabbing, when five aphids per condition were removed from the plants, frozen in liquid nitrogen, and stored at -80°C prior to RNA extraction. Unstabbed control aphids, raised and treated under similar conditions, were frozen as well. In addition, 20 stabbed and 20 unstabbed aphids were monitored over a period of 96 hours for survival (Figure S3c, below). Though it is possible that responses to stabs with the virus-slurry would be a result of exposure to other pathogens in the ground-up cadaver, stabs with slurry from non-virus killed aphid cadavers do not lead to a significant increase in mortality (Parker, unpublished data), and thus we expect that responses can be largely attributed to virus exposure.

Stress experiment. For the fifth experiment, we exposed aphids to a number of stressors. We exposed one group of aphids on plants to 3 μL of (*E*)-β farnesene (EBF) (1 μg/μL in hexane), which we placed on a small piece of filter paper near the base of each plant. We enclosed the plants under a plastic cover with no ventilation. Within minutes of exposure, we froze five EBF-exposed aphids, and five control, unexposed aphids. Twelve hours later, we froze another five aphids that had been exposed to EBF. We heat shocked a second group of aphids following standard procedures [7]. After two hours at 36^{o}C, we froze five heat-shocked aphids and then maintained another five heat-shocked aphids at 20^{o}C for another eight hours prior to freezing. To assess the affects of starvation, we placed a third batch of aphids into sterile Petri dishes with moistened filter paper for 12 hours prior to freezing.

Quantitative PCR. For each sample, we extracted RNA from five whole aphids using a Qiagen RNA easy tissue kit and prepared cDNA from each sample using a Qiagen Quantitect reverse transcription kit. We carried out expression studies, utilizing the delta-delta CT method, on an Applied Biosystems Step One Plus machine. Each reaction contained 10 μL AB Power SYBR PCR master mix, 300 nM of each primer, approximately 100 ng cDNA, and water to a total volume of 20 μL. We designed primers utilizing Primer3 [8] or Primer Express (Applied Biosystems) to amplify approximately 100 bps of the gene of interest (Table S3, below). Primers spanned an exon-exon boundary where possible. For each sample, we carried out three separate reactions for each primer pair, and averaged the comparative threshold cycle (Ct) among the three values. We standardized all Ct values for the gene of interest relative to the Ct values for the endogenous control gene actin, yielding the delta Ct value. We then standardized relative to the appropriate control, yielding the delta-delta Ct value. Finally, these delta-delta Ct values were standardized such that each control treatment average was one, yielding the relative quantity (RQ) values, to allow for comparison with studies reporting fold-changes.

Table S1. Pea aphid immune and stress gene list. Genes are listed in the approximate order to which they are mentioned in the text. The last column indicates results of a tblastn search (contig ID and e-value) of all identified pea aphid genes against an EST sequence database for *Myzus persicae*, the green peach aphid. All pea aphid sequences are accessible at AphidBase Gbrowse [9]. All green peach aphid sequences can be downloaded at AphidBase Downloads (files Myzus454 and MyzusSanger) [10].

Class	Role in Insect Immunity	Gene Symbol	Gene Name	Other Gene Names	ACYPI identifier	NCBI gene ID	NCBI transcript ID	NCBI protein ID	*Myzus* ID (e-value)
recognition	bacterial recognition	PGRP	Peptidoglycan recognition protein		not found	not found	not found	not found	n/a
recognition	bacterial and fungal pattern recognition	GNBP1	Gram Negative Binding Protein 1		ACYPI005376	100164352	XM_001944438.1	XP_001944473.1	6372 (7e-134)
recognition	bacterial and fungal pattern recognition	GNBP2	Gram Negative Binding Protein 2		ACYPI006143	100165182	XM_001947795.1	XP_001947830.1	3883 (4e-178)
recognition	bacterial recognition, induction of phenoloxidase	Ctl1	C-type Lectin 1		ACYPI004676	100163601	XM_001944997.1	XP_001945032.1	26603 (3e-39)
recognition	bacterial recognition, induction of phenoloxidase	Ctl2	C-type Lectin 2		ACYPI005998	100165024	XM_001946121.1	XP_001946156.1	26603 (3e-38)
recognition	bacterial recognition, induction of phenoloxidase	Ctl3	C-type Lectin 3		ACYPI004682	100163607	XM_001943575.1	XP_001943610.1	60035 (3e-6)
recognition	bacterial recognition, induction of phenoloxidase	Ctl	C-type Lectin galactose-binding		ACYPI009411	100168734	XM_001951375.1	XP_001951410.1	3644 (3e-130)
recognition	bacterial recognition, induction of phenoloxidase	Ctl	C-type Lectin selectin-like		ACYPI003045	100161853	XM_001942942.1	XP_001942977.1	73637 (1e-31)
recognition	several roles have been hypothesized	gale1	galectin 1	Galactoside-binding soluble lectin	ACYPI001371	100160038	XM_001943734.1	XP_001943769.1	30445 (2e-6)
recognition	several roles have been hypothesized	gale2	galectin 2	Galactoside-binding soluble lectin	ACYPI000409	100158995	XM_001947578.1	XP_001947613.1	24692 (8e-32)
recognition	bind to lipoproteins, bacteria	sr-CI	Scavenger receptor class C, type I		not found	not found	not found	not found	n/a
recognition	bind to lipoproteins, bacteria	sr-CII	Scavenger receptor class C, type II		not found	not found	not found	not found	n/a
recognition	bind to lipoproteins, bacteria	sr-CIII	Scavenger receptor class C, type III		not found	not found	not found	not found	n/a
recognition	bind to lipoproteins, bacteria	sr-CIV	Scavenger receptor class C, type IV		not found	not found	not found	not found	n/a
recognition	receptor in phagocytosis and microbial binding	eater	eater		not found	not found	not found	not found	n/a
recognition	receptor in phagocytosis and microbial binding	nim-C1	nimrod C1		not found	not found	not found	not found	n/a
signaling	toll pathway	spz1-1	spätzle 1Bi	spatzle 1b	ACYPI004362	100163265	XM_001950373.1	XP_001950408.1	649 (1e-99)
signaling	toll pathway	spz1-2	spätzle 1Bii	spatzle 1b	ACYPI001858	100160571	XM_001947931.1	XP_001947966.1	649 (6e-119)
signaling	toll pathway	spz1-3	spätzle 1-3	spatzle 1b, spaetzle 1b	ACYPI41073				649 (4e-83)
signaling	toll pathway	spz1-4	spätzle 1-4	spatzle 1b, spaetzle 1b	ACYPI52992				649 (5e-71)

Class	Role in Insect Immunity	Gene Symbol	Gene Name	Other Gene Names	ACYPI identifier	NCBI gene ID	NCBI transcript ID	NCBI protein ID	*Myzus* ID (e-value)
signaling	toll pathway	spz1-5	spätzle 1-5	spatzle 1b, spaetzle 1b	ACYPI21155				649 (3e-65)
signaling	toll pathway	spz2	spätzle 2	spatzle 2	ACYPI003414	100162252	XM_001948424.1	XP_001948459.1	649 (2e-10)
signaling	toll pathway	spz3	spätzle 3	spatzle 3	ACYPI55738				649 (3e-4)
signaling	toll pathway	spz4	spätzle 4	spatzle 4	ACYPI006811	100165897	XM_001949337.1	XP_001949372.1	no hit
signaling	toll pathway	spz6	spätzle 6	spatzle 6	ACYPI001990	100160712	XM_001944011.1	XP_001944046.1	no hit
signaling	toll pathway	spz-like	spätzle-like, partial	spatzle, spaetzle	ACYPI009165	100168467	XM_001946778.1	XP_001946813.1	no hit
signaling	some tolls function in toll signaling pathway	18w	18 wheeler	Toll, Toll-2	ACYPI008698	100167952	XM_001946908.1	XP_001946943.1	6369 (6e-28)
signaling	some tolls function in toll signaling pathway	Toll	Toll-like		ACYPI000177	100158739	XM_001942698.1	XP_001942733.1	5712 (4e-115)
signaling	some tolls function in toll signaling pathway	Toll	Toll-like		ACYPI002340	100161089	XM_001946411.1	XP_001946446.1	5712 (4e-122)
signaling	some tolls function in toll signaling pathway	Toll	Toll-like, partial		ACYPI004287	100163187	XM_001949147.1	XP_001949182.1	5712 (2e-109)
signaling	some tolls function in toll signaling pathway	Toll-6	Toll-6		ACYPI005417	100164395	XM_001947289.1	XP_001947324.1	6369 (1e-34)
signaling	some tolls function in toll signaling pathway	Toll	Toll-like		ACYPI008268	100167471	XM_001950727.1	XP_001950762.1	6369 (5e-19)
signaling	some tolls function in toll signaling pathway	Tollo	Tollo	Toll-8	ACYPI002754	100161538	XM_001948531.1	XP_001948566.1	6369 (1e-44)
signaling	toll pathway	tub	tube	interleukin-1 receptor-associated kinase 4	ACYPI006580	100165647	XM_001950581.1	XP_001950616.1	4836 (2e-113)
signaling	toll pathway	Myd88	myeloid differentiation primary response gene		ACYPI001638	100160335	XM_001948285.1	XP_001948320.1	1295 (9e-168)
signaling	toll pathway	pll	pelle		ACYPI009928	100169297	XM_001942995.1	XP_001943030.1	4836 (2e-38)
signaling	toll pathway	cact	cactus		ACYPI006820	100165906	XM_001950793.1	XP_001950828.1	259 (3e-135)
signaling	toll pathway	cactin	cactin		ACYPI006968	100166064	XM_001952252.1	XP_001952287.1	2493 (0)
signaling	toll pathway	Pli	Pellino		ACYPI001694	100160395	XM_001946247.1	XP_001946282.1	9336 (3e-71)
signaling	toll pathway	Traf1	TNF-receptor-associated factor 1		ACYPI000855	100159489	XM_001948320.1	XP_001948355.1	24857 (2e-7)
signaling	toll pathway	Traf2	TNF-receptor-associated factor 2	dTraf2, Traf6, TNF-receptor-associated factor 6	not found	not found	not found	not found	n/a
signaling	toll pathway	dl	dorsal		ACYPI003588	100162436	XM_001947394.1	XP_001947429.1	6239 (2e-119)
signaling	toll pathway	dlB	dorsal B		ACYPI005133	100164092	XM_001949463.1	XP_001949498.1	6239 (3e-124)
signaling	jak/stat pathway	dome-1	domeless 1		ACYPI21995	100294629			57862 (9e-8)

Class	Role in Insect Immunity	Gene Symbol	Gene Name	Other Gene Names	ACYPI identifier	NCBI gene ID	NCBI transcript ID	NCBI protein ID	*Myzus* ID (e-value)
signaling	jak/stat pathway	dome-2	domeless 2		ACYPI004970	100163919			57862 (6e-8)
signaling	jak/stat pathway	dome-3	domeless 3		ACYPI21996	100294630			57862 (3e-8)
signaling	jak/stat pathway	dome-4	domeless 4, 5' partial		ACYPI40957	100294721			no hit
signaling	jak/stat pathway	Jak	Janus kinase	hopscotch	ACYPI008118	100167312	XM_001948012.1	XP_001948047.1	1739 (2e-158)
signaling	jak/stat pathway	Stat92E-1	Signal-transducer and activator of transcription 1	Stat	ACYPI002351	100161101	XM_001946610.1	XP_001946645.1	2173 (4e-114)
signaling	jak/stat pathway	Stat92E-2	Signal-transducer and activator of transcription 1 , partial	Stat	ACYPI005642	100164649	XM_001943623.1	XP_001943658.1	65719 (5e-27)
signaling	jak/stat pathway	upd	unpaired		not found	not found	not found	not found	n/a
signaling	imd pathway	imd	immune deficiency		not found	not found	not found	not found	n/a
signaling	imd pathway	dFadd	dFadd	*Drosophila BG4, FADD*	not found	not found	not found	not found	n/a
signaling	imd pathway	Dredd	Death related ced-3	caspase 8	not found	not found	not found	not found	n/a
signaling	imd pathway	Rel	Relish	Nf-KB, REL	not found	not found	not found	not found	n/a
signaling	imd pathway	Tab2	TAK1-associated Binding Protein2		ACYPI002796	100161584	XM_001950344.1	XP_001950379.1	5307 (5e-106)
signaling	imd pathway	Tak1	TGF-β activated kinase 1		ACYPI001063	100159713	XM_001944422.1	XP_001944457.1	2842 (3e-88)
signaling	imd pathway	key	kenny	IKKgamma	not found	not found	not found	not found	not found
signaling	imd pathway	Iap2	Inhibitor of apoptosis 2		ACYPI000445	100159034	XM_001942899.1	XP_001942934.1	60268 (2e-27)
signaling	imd pathway	ird5	immune response deficiency 5	IKK	ACYPI004933	100163880	XM_001952347.1	XP_001952382.1	7578 (3e-90)
signaling	jnk pathway	hep	hemipterous		ACYPI005993	100165019	XM_001944492.1	XP_001944527.1	8122 (8e-46)
signaling	jnk pathway	bsk	basket	MAPK, JNK	ACYPI004372	100163276	XM_001945425.1	XP_001945460.1	1929 (1e-88)
signaling	jnk pathway	Jra	Jun-related antigen		ACYPI002386	100161138	XM_001947521.1	XP_001947556.1	6809 (1e-89)
signaling	jnk pathway	kay	kayak		not found	not found	not found	not found	n/a
signaling	JNK pathway	egr	eiger		ACYPI001133	100159786	XM_001952555.1	XP_001952590.1	1287(2e-131)
response	antimicrobial peptide		Abaecin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Alloferon		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Andropin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Apisimin		not found	not found	not found	not found	n/a

Class	Role in Insect Immunity	Gene Symbol	Gene Name	Other Gene Names	ACYPI identifier	NCBI gene ID	NCBI transcript ID	NCBI protein ID	*Myzus* ID (e-value)
response	antimicrobial peptide	Att	Attacin		not found	not found	not found	not found	n/a
response	antimicrobial peptide	Cec	Cecropin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Coleoptericin		not found	not found	not found	not found	n/a
response	antimicrobial peptide	Def	Defensin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Diptericin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Drosocin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Drosomycin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Formicin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Gambicin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Gomesin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Heliocin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Holotricin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Lebocin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Megourin		not found, though see text	not found	not found	not found	n/a
response	antimicrobial peptide		Metchnikowin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Moricin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Penaeidin		not found, though see text	not found	not found	not found	n/a
response	antimicrobial peptide		Polyphemusin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Spingerin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Tachyplesin		not found	not found	not found	not found	n/a
response	antimicrobial peptide		Virescein		not found	not found	not found	not found	n/a
response	antimicrobial peptide	Thm1	Thaumatin1	Tha, Thn	ACYPI005841	100164856	XM_001942683.1	XP_001942718.1	81361 (1e-8)
response	antimicrobial peptide	Thm2	Thaumatin2	Tha, Thn	ACYPI001394	100160062	XM_001942537.1	XP_001942572.1	81361 (2e-8)
response	antimicrobial peptide	Thm3	Thaumatin3	Tha, Thn	ACYPI009605	100168942	XM_001942744.1	XP_001942779.1	81361 (2e-9)
response	antimicrobial peptide	Thm4	Thaumatin4	Tha, Thn	ACYPI003287	100162111	XM_001942495.1	XP_001942530.1	81361 (2e-9)
response	antimicrobial peptide	Thm5	Thaumatin5	Tha, Thn	ACYPI007568	100166717	XM_001951871.1	XP_001951906.1	9364 (5e-22)

Class	Role in Insect Immunity	Gene Symbol	Gene Name	Other Gene Names	ACYPI identifier	NCBI gene ID	NCBI transcript ID	NCBI protein ID	*Myzus* ID (e-value)
response	antimicrobial peptide	Thm6	Thaumatin6	Tha, Thn	ACYPI005301	100164271	XM_001942753.1	XP_001942788.1	81361 (3e-15)
response	microbial degradation	Lys1	Lysozyme, i-type		ACYPI002175	100160909	XM_001949053.1	XP_001949088.1	3094 (8e-63)
response	microbial degradation	Lys2	Lysozyme, i-type		ACYPI009125	100168424	XM_001949177.1	XP_001949212.1	584 (3e-80)
response	microbial degradation	Lys3	Lysozyme, i-type		ACYPI008509	100167742	XM_001949283.1	XP_001949318.1	74639 (7e-27)
response	fungal degradation	Cht1	Chitinase-like protein 1		ACYPI001365	100160032	XM_001943565.1	XP_001943600.1	3640 (0)
response	fungal degradation	Cht2	Chitinase-like protein 2	Drosophila chitinase 3	ACYPI010095	100169480	XM_001943003.1	XP_001943038.1	5202 (1e-52)
response	fungal degradation	Cht3	Chitinase-like protein 3		ACYPI001396	100160065	XM_001942561.1	XP_001942596.1	5202 (6e-73)
response	fungal degradation	Cht4	Chitinase-like protein 4		ACYPI006403	100165452	XM_001950345.1	XP_001950380.1	5202 (2e-168)
response	fungal degradation	Cht5	Chitinase-like protein 5		ACYPI009964	100169337	XM_001947381.1	XP_001947416.1	5202 (2e-52)
response	fungal degradation	Cht6	Chitinase-like protein 6		ACYPI009878	100169240	XM_001952683.1	XP_001952718.1	5202 (3e-50)
response	fungal degradation	Cht7	Chitinase-like protein 7		ACYPI005756	100164767	XM_001947852.1	XP_001947887.1	1327 (4e-162)
response	general stress response	TotA	Turandot A		not found	not found	not found	not found	n/a
response	general stress response	TotB	Turandot B		not found	not found	not found	not found	n/a
response	general stress response	TotC	Turandot C		not found	not found	not found	not found	n/a
response	general stress response	TotF	Turandot F		not found	not found	not found	not found	n/a
response	general stress response	TotM	Turandot M		not found	not found	not found	not found	n/a
response	general stress response	TotZ	Turandot Z		not found	not found	not found	not found	n/a
response	mark pathogens for phagocytosis	TepI	Thiolester containing protein I		not found	not found	not found	not found	n/a
response	mark pathogens for phagocytosis	TepII	Thiolester containing protein II		not found	not found	not found	not found	n/a
response	mark pathogens for phagocytosis	TepIII-1	Thiolester containing protein III – 1, partial		ACYPI005292	100164261	XM_001944313.1	XP_001944348.1	no hit
response	mark pathogens for phagocytosis	TepIII-2	Thiolester containing protein III – 2		ACYPI000145	100158705	XM_001945685.1	XP_001945720.1	no hit
response	mark pathogens for phagocytosis	TepIV	Thiolester containing protein IV		not found	not found	not found	not found	n/a
response	prophenoloxidase response	ProPO1	Prophenoloxidase 1	Diphenol oxidase A3	ACYPI001367	100160034	XM_001949272.1	XP_001949307.1	5431 (5e-147)
response	prophenoloxidase response	ProPO2	Prophenoloxidase 2	Diphenol oxidase A3	ACYPI004484	100163393	XM_001951102.1	XP_001951137.1	5721 (e3-133)
response	production of nitric oxide, a toxic gas	Nos	Nitric oxide synthase		ACYPI001689	100160390	XM_001946174.1	XP_001946209.1	462 (1e-35)

Class	Role in Insect Immunity	Gene Symbol	Gene Name	Other Gene Names	ACYPI identifier	NCBI gene ID	NCBI transcript ID	NCBI protein ID	*Myzus* ID (e-value)
response	general stress response	Hsc5	Heat shock cognate 5		ACYPI004693	100163620	XM_001950464.1	XP_001950499.1	
response	general stress response	Hsc70	Heat shock cognate 70		ACYPI007166	100166283	XM_001948031.1	XP_001948066.1	1624 (0)
response	general stress response	Hsc70-1	Heat shock cognate 70 - 1		ACYPI000474	100159065	XM_001951198.1	XP_001951233.1	182 (0)
response	general stress response	Hsc70-2	Heat shock cognate 70 - 2		ACYPI004809	100163748	XM_001951351.1	XP_001951386.1	182 (0)
response	general stress response	Hsc70Cb	Heat shock cognate 70 - Cb		ACYPI004544	100163455	XM_001951757.1	XP_001951792.1	2297 (0)
response	general stress response	Hsp14	Heat shock protein 14		ACYPI002719	100161502	XM_001945733.1	XP_001945768.1	3375 (0)
response	general stress response	Hsp21.4	Heat shock protein 21.4		ACYPI003907	100162777	XM_001949367.1	XP_001949402.1	6174 (2e-108)
response	general stress response	Hsp60	Heat shock protein 60		ACYPI009253	100168563	XM_001951338.1	XP_001951373.1	317 (6e-155)
response	general stress response	Hsp70Aa	Heat shock protein 70Aa		ACYPI009117	100168413	XM_001951880.1	XP_001951915.1	182 (0)
response	general stress response	Hsp70Ab	Heat shock protein 70Ab		ACYPI007961	100167145	XM_001949626.1	XP_001949661.1	182 (0)
response	general stress response	Hsp70Ba	Heat shock protein 70Ba		ACYPI008763	100168026	XM_001949802.1	XP_001949837.1	182 (0)
response	general stress response	Hsp83	Heat shock protein 83		ACYPI009380	100168702	XM_001944726.1	XP_001944761.1	1268 (0)
response	general stress response	Hsp83	Heat shock protein 83		ACYPI002010	100160736	XM_001943137.1	XP_001943172.1	1268 (0)
response	general stress response	Hsp90	Heat shock protein 90		ACYPI002398	100161155	XM_001951175.1	XP_001951210.1	5927 (1e-149)
response	general stress response	Hsp90	Heat shock protein 90		ACYPI009915	100169283	XM_001948902.1	XP_001948937.1	3803 (0)
response	detoxification	Gst	Glutathione S-transferase-like, microsomal		ACYPI004835	100163775	XM_001951200.1	XP_001951235.1	3849 (8e-76)
response	detoxification	Gst	Glutathione S-transferase-like, microsomal		ACYPI006691	100165764	XM_001946296.1	XP_001946331.1	3849 (5e-37)
response	detoxification	Gst	Glutathione S-transferase-like, sigma class		ACYPI000794	100159421	XM_001952064.1	XP_001952099.1	3513 (3e-99)
response	detoxification	Gst	Glutathione S-transferase-like, sigma class		ACYPI002127	100160859	XM_001952005.1	XP_001952040.1	3896 (1e-105)
response	detoxification	Gst	Glutathione S-transferase-like, sigma class		ACYPI002679	100161459	XM_001952021.1	XP_001952056.1	3896 (1e-100)
response	detoxification	Gst	Glutathione S-transferase-like, sigma class		ACYPI009326	100168645	XM_001952392.1	XP_001952427.1	1197 (6e-82)
response	detoxification	Gst	Glutathione S-transferase-like, sigma class		ACYPI009519	100168850	XM_001946569.1	XP_001946604.1	1197 (2e-102)
response	detoxification	Gst	Glutathione S-transferase-like, sigma class		ACYPI009520	100168850	XM_001946504.1	XP_001946539.1	1197 (2e-102)
response	detoxification	GstD	Glutathione S-transferase, delta class		ACYPI006899	100165990	XM_001951401.1	XP_001951436.1	2376 (4e-56)
response	detoxification	GstD	Glutathione S-transferase, delta class		ACYPI009586	100168923	XM_001952561.1	XP_001952596.1	2376 (5e-38)

Class	Role in Insect Immunity	Gene Symbol	Gene Name	Other Gene Names	ACYPI identifier	NCBI gene ID	NCBI transcript ID	NCBI protein ID	*Myzus* ID (e-value)
response	detoxification	GstD10	Glutathione S-transferase, delta class		ACYPI008042	100167231	XM_001948159.1	XP_001948194.1	2376 (7e-73)
response	detoxification	GstD4	Glutathione S-transferase, delta class		ACYPI001068	100159718	XM_001942679.1	XP_001942714.1	2376 (1e-84)
response	detoxification	GstD6	Glutathione S-transferase, delta class		ACYPI006598	100165666	XM_001952338.1	XP_001952373.1	2376 (4e-53)
response	detoxification	GstD6	Glutathione S-transferase, delta class, partial		ACYPI008550	100167788	XM_001952381.1	XP_001952416.1	2376 (3e-52)
response	detoxification	GstD8	Glutathione S-transferase, delta class		ACYPI008657	100167906	XM_001942576.1	XP_001942611.1	2376 (9e-112)
response	detoxification	GstD9	Glutathione S-transferase, delta class		ACYPI005620	100164626	XM_001950500.1	XP_001950535.1	569 (1e-123)
response	detoxification	Gst	Glutathione S-transferase, theta class		ACYPI007233	100166353	XM_001949321.1	XP_001949356.1	6981 (2e-79)
response	detoxification	Gst	Glutathione S-transferase, theta class		ACYPI009122	100168419	XM_001949359.1	XP_001949394.1	9615 (3e-73)
response	alarm pheromone production	IPPS	Isoprenyl diphosphate synthase		ACYPI000050	100144905	NM_001126161.3	NP_001119633.3	912 (0)
response	alarm pheromone production	FPPS	similar to Farnesyl diphosphate synthase 2		ACYPI007080	100166187	XM_001950388.1	XP_001950423.1	912 (0)

Table S2. Samples for qPCR expression study.

Sample Name	Sample Handling Notes
Bacterial Stabbing Experiment	
No stab control	Frozen 8hrs after exposure of treatment aphids
Sterile stab	Frozen 8hrs after exposure
E.coli stab	Frozen 8hrs after exposure
Gram- pathogen stab	Frozen 8hrs after exposure
Gram+ pathogen stab	Frozen 8hrs after exposure
Bacterial Feeding Experiment	
Feed control	Frozen 12hrs after exposure to diet
E. coli feed	Frozen 12hrs after exposure to diet
Gram- pathogen feed	Frozen 12hrs after exposure to diet
Fungal Shower Experiment	
Fungus control	Frozen 24hrs after exposure
Fungus infected	Frozen 24hrs after exposure
Viral Stabbing Experiment	
Virus control (no stab)	Frozen 16hrs after exposure
Virus infected	Frozen 16hrs after exposure
Stress Experiment	
No stress Control	Frozen at beginning of stress experiment
Alarm pheromone	Frozen minutes after exposure to EBF
Post alarm pheromone	Frozen 12hrs after exposure to EBF
Heat stress	Frozen immediately after 36^{o}C heat shock
Post heat stress	Frozen 8hrs after being returned to 20^{o}C
Starvation	Frozen 12hrs after being removed from plant

Table S3. Primers for qPCR expression study.

Gene symbol	Gene name	Putative function	ACYPI ID	Primer Pair (5' to 3')
GNBP2	gram-negative binding protein 2	recognition	ACYPI006143	gnbp2_1f: AATTTCCGTGATGGGTGTTTAAGT gnbp2_1r: TTTGTTTTCATTCCATGTTGATGAC
Gale1	galectin 1	recognition and response	ACYPI001371	gale_1f: GCTCCAATACTCAATCCGACTCTT gale_1r: CATCGTCCTTGTTTCAAACC
Tl	toll	toll signaling pathway	ACYPI000177	tl1_1f: GAGCTCACCGTTTAAACTTTGTCA tl1_1r: CATCAACTGAACGAGCAATTTGA
Cact	cactus	toll signaling pathway	ACYPI006820	cact_1f: GATGGCCAAAGTGCTCTTCATT cact_1r: GAGCTTTCATTAGGTGTTTCACAATTT
Dl	dorsal 1b	toll signaling pathway	ACYPI003588	dl_1f: CAAGAGAATAGAAAACCACATCGTCTA dl_1r: AAACATCAGTGTTATGCGGCTAAG
DlB	dorsal 1b	toll signaling pathway	ACYPI005133	dlb_1f: CTCTCAGAGTACGAGAAGAAATAAGAGTAGAT dlb_1r: AAACATCAGTGTTATGCGGCTAAG
Myd88	myeloid differentiation primary response gene	toll signaling pathway	ACYPI001638	myd88_1f: TGCATGTTAAATGCCACGAAA myd88_1r: TCCTCTCCAATCCCTGGGTAA
IRD5	immune response deficient 5	imd/jnk pathway	ACYPI004933	IRD5_1f: TGCTTATCTTGCACCGGAAGT IRD5_1r: ACTATGACTCCAACACTCCACATATCTAA
Bsk	basket (JNK)	jnk pathway	ACYPI004372	Basket_1f: TTTGATCGATTATTCCCTGATGTACT Basket_1r: GTGCTTGGCTTGCTTTTAGTTTATT
JRA1	Jun-related antigen	imd/jnk pathway	ACYPI002386	Jra_1f: AAATCAAACTCGAAAGGAAAAGACA Jra_1r: TTCGGCGGCATTTGGA
IAP2	inhibitor of apoptosis 2	imd/jnk pathway	ACYPI000445	Iap2_10f: TCGATGAACACAAACGTCACAA Iap2_10r: GTTCACCAGTTTCCTTATGATTTTCAA
Stat92E-2	Signal-transducer and activator of transcription 2	Jak/Stat pathway	ACYPI005642	Stat2_37f: TCATTAGTTCAGTGGAAACGTCAAC Stat2_37r: AATCACACAATTTCTCACACCAAGTT
Lys1	lysozyme, i-type	response to bacteria	ACYPI002175	lysoz1_1f: CGCACAGGACTGCAACCA lysoz1_1r: GGATGGCCGCGTAATCAG
Lys2	lysozyme, i-type	response to bacteria	ACYPI009125	Lys2_10f: GCGTCAGAGACCCGTATTGC Lys2_10r: GCAATCCTTTGCGTATCTCTGTATG
Lys3	lysozyme, i-type	response to bacteria	ACYPI008509	lysoz3_1f: CCGGTCAGTAGCAGAGGAAAGT lysoz3_1r: ATGAGCTCTCGCGTAGTTTGG
Thm2	thaumatin 2	response (antimicrobial peptide)	ACYPI001394	Tha2_r1f: CAACAGTAAAGGAAAATGCGAAAC Tha2_r1r: TGGCACGCCCATGATACC
Thm3	thaumatin 3	response (antimicrobial peptide)	ACYPI009605	Tha3_1f: GGGCAGGCAGGATTTGG Tha3_1r: TTGGATCTTGTTCCCGCAAT
Thm4	thaumatin 4	response (antimicrobial peptide)	ACYPI003287	Tha4_1f: GGCGGGCAGGATTTGG Tha4_1r: GTGGATCTTGTTCCCGCAAT
Thm6	thaumatin 6	response (antimicrobial peptide)	ACYPI005301	Tha6_1f: AAAATGCAGCGCTCAAGGA Tha6_1r: CACTTGATCTTGTTCCCGCAAT
HSP60	heat shock protein 60	response to stress	ACYPI009253	hsp60_1f: GATGCAATGAACGACGAATATGTTA hsp60_1r: CTGACAACTTTGGTTGGATCGA
HSC70	heat shock cognate 70	response to stress	ACYPI007166	HP3f: TTGGGTGGAGAAGATTTTGA HP3r: ATGACTGGCAGAAAGACCAC
HSP83	heat shock protein 83	response to stress	ACYPI002010	HP6f: CCGTACTGATCCTGGTGAAC HP6r: GCCAATGAATTGAGAGTGCT

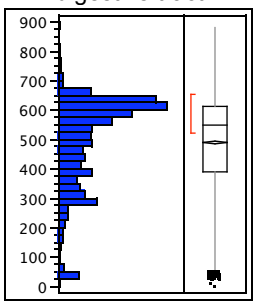
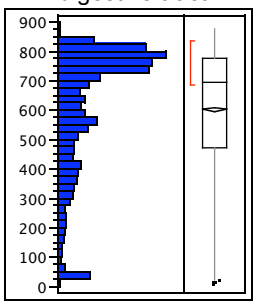
Table S4. Expression of recognition and signaling genes. Values indicate the relative expression of a gene in a sample relative to expression in the appropriate control sample, +/- one standard deviation. For the bacterial stabbing experiment, we compared a sterile stab sample to a no stab sample only for genes that showed greater than 2-fold upregulation in one of the bacteria-stabbed samples. Relative expression values should be interpreted with caution as they are based on only a single experimental replicate pooling five aphids.

	GNBP2	GALE1	TOLL	CACTUS	DOR1	DOR1B	MYD88	IRD5	JNK	JRA1	IAP2	JAK	STAT92E2
No stab control	1.00+/-1.05	1.00+/-0.08	1.00+/-0.03	1.00+/-0.06	1.00+/-0.16	1.00+/-0.05	1.00+/-0.12	1.00+/-0.07	1.00+/-0.06	1.00+/-0.02	1.00+/-0.04	1.00+/-0.12	1.00+/-0.11
Sterile stab			3.01+/-0.07	0.66+/-0.02							0.48+/-0.03		
E.coli stab	0.49+/-0.05	0.66+/-0.02	5.53+/-0.48	2.18+/-0.13	1.66+/-0.07	0.72+/-0.03	0.82+/-0.06	0.45+/-0.04	0.87+/-0.05	1.19+/-0/03	1.64+/-0.08	0.84+/-0.78	0.46+/-0.04
Gram-pathogen stab	0.59+/-0.06	0.78+/-0.12	2.52+/-0.21	1.78+/-0.33	0.84+/-0.06	0.65+/-0.02	1.26+/-0.08	0.43+/-0.06	0.76+/-0.02	0.85+/-0.06	2.00+/-0.06	0.77+/-0.06	0.59+/-0.04
Gram+pathogen stab	0.82+/-0.04	0.69+/-0.02	2.28+/-0.22	1.42+/-0.06	0.89+/-0.05	0.51+/-0.05	0.99+/-0.05	0.51+/-0.05	0.69+/-0.04	0.91+/-0.05	2.43+/-0.16	0.73+/-0.04	0.42+/-0.02
Feed control	1.00+/-0.05	1.00+/-0.07	1.00+/-0.10	1.00+/-0.07	1.00+/-0.05	1.00+/-0.08	1.00+/-0.07	1.00+/-0.03	1.00+/-0.10	1.00+/-0.07	1.00+/-0.06	1.00+/-0.08	1.00+/-0.08
E. coli feed	1.01+/-0.05	0.77+/-0.05	1.45+/-0.19	2.06+/-0.12	0.88+/-0.07	0.80+/-0.05	0.91+/-0.03	1.00+/-0.03	0.72+/-0.05	1.04+/-0.07	1.15+/-0.06	0.77+/-0.02	0.74+/-0.10
Gram-pathogen feed	0.82+/-0.06	0.63+/-0.05	1.41+/-0.15	1.35+/-0.05	0.84+/-0.08	0.70+/-0.03	0.63+/-0.03	1.03+/-0.04	0.70+/-0.05	0.95+/-0.03	1.46+/-0.10	0.64+/-0.02	0.69+/-0.07
Fungus control	1.00+/-0.07	1.00+/-0.03	1.00+/-0.08	1.00+/-0.02	1.00+/-0.07	1.00+/-0.04	1.00+/-0.09	1.00+/-0.09	1.00+/-0.06	1.00+/-0.08	1.00+/-0.04	1.00+/-0.07	1.00+/-0.04
Fungus infected	0.91+/-0.03	1.20+/-0.18	1.16+/-0.10	0.87+/-0.11	1.06+/-0.02	1.00+/-0.07	1.00+/-0.05	0.62+/-0.07	1.30+/-0.10	1.00+/-0.03	0.88+/-0.03	1.32+/-0.07	1.32+/-0.09
Virus control	1.00+/-0.02	1.00+/-0.04	1.00+/-0.03	1.00+/-0.03	1.00+/-0.08	1.00+/-0.06	1.00+/-0.02	1.00+/-0.04	1.00+/-0.06	1.00+/-0.07	1.00+/-0.10	1.00+/-0.03	1.00+/-0.12
Virus infected	0.88+/-0.03	1.01+/-0.04	1.23+/-0.14	0.83+/-0.02	1.59+/-0.05	0.97+/-0.04	0.87+/-0.05	0.92+/-0.06	0.93+/-0.03	1.16+/-0.06	0.98+/-0.07	1.40+/-0.11	2.42+/-0.27
No stress control	1.00+/-0.10	1.00+/-0.14	1.00+/-0.03	1.00+/-0.02	1.00+/-0.07	1.00+/-0.07	1.00+/-0.04	1.00+/-0.06	1.00+/-0.04	1.00+/-0.04	1.00+/-0.08	1.00+/-0.06	1.00+/-0.06
Alarm pheromone	0.79+/-0.05	0.80+/-0.07	0.94+/-0.03	2.41+/-0.24	0.96+/-0.05	0.87+/-0.04	0.74+/-0.02	0.95+/-0.13	0.97+/-0.06	0.96+/-0.03	1.37+/-0.09	1.15+/-0.07	0.98+/-0.05
Post alarm pheromone	0.54+/-0.03	0.94+/-0.12	0.34+/-0.05	2.46+/-0.10	0.78+/-0.15	0.98+/-0.07	0.96+/-0.07	1.03+/-0.14	0.99+/-0.06	0.64+/-0.13	2.11+/-0.14	1.08+/-0.08	1.00+/-0.07
Heat stress	1.35+/-0.07	1.14+/-0.05	0.74+/-0.08	1.48+/-0.06	2.59+/-0.17	1.71+/-0.06	1.01+/-0.03	0.94+/-0.07	1.14+/-0.07	1.17+/-0.16	0.68+/-0.02	1.27+/-0.22	1.97+/-0.12
Post heat stress	0.76+/-0.04	0.69+/-0.09	0.48+/-0.02	1.68+/-0.10	0.75+/-0.11	0.63+/-0.08	0.58+/-0.03	0.67+/-0.04	0.77+/-0.01	0.69+/-0.02	2.11+/-0.12	0.76+/-0.02	0.58+/-0.07
Starvation	2.14+/-0.42	1.16+/-0.03	0.64+/-0.05	3.29+/-0.27	1.79+/-0.12	1.36+/-0.06	0.76+/-0.06	0.86+/-0.11	1.20+/-0.06	0.44+/-0.08	1.13+/-0.14	1.37+/-0.07	1.03+/-0.12

Table S5. Expression of response genes. Values indicate the relative expression of a gene in a sample relative to expression in the appropriate control sample, +/- S.D. For the bacterial stabbing experiment, we compared a sterile stab sample to a no stab sample only for genes that showed greater than 2-fold upregulation in one of the bacteria-stabbed samples. Relative expression values should be interpreted with caution as they are based on only a single experimental replicate pooling five aphids.

	LYS1	LYS2	LYS3	THM2	THM3	THM4	THM6	HSP60	HSC70	HSP83
No stab control	1.00+/-0.05	1.00+/-0.07	1.00+/-0.03	1.00+/-0.04	1.00+/-0.09	1.00+/-0.06	1.00+/-0.06	1.00+/-0.05	1.00+/-0.05	1.00+/-0.05
Sterile stab				0.74+/-0.07						
E.coli stab	1.00+/-0.03	1.35+/-0.08	1.76+/-0.15	2.37+/-0.07	0.34+/-0.05	0.29+/-0.07	0.39+/-0.03	0.81+/-0.04	0.58+/-0.04	0.80+/-0.06
Gram-pathogen stab	1.12+/-0.07	1.41+/-0.05	1.57+/-0.07	1.80+/-0.03	1.06+/-0.04	0.86+/-0.08	1.60+/-0.06	0.96+/-0.05	0.69+/-0.05	1.26+/-0.07
Gram+pathogen stab	1.12+/-0.04	1.59+/-0.17	1.62+/-0.06	0.56+/-0.05	0.58+/-0.04	0.51+/-0.04	1.01+/-0.07	0.84+/-0.04	0.66+/-0.02	0.97+/-0.04
Feed control	1.00+/-0.02	1.00+/-0.12	1.00+/-0.04	1.00+/-0.04	1.00+/-0.05	1.00+/-0.05	1.00+/-0.06	1.00+/-0.09	1.00+/-0.04	1.00+/-0.07
E. coli feed	2.16+/-0.13	1.17+/-0.05	1.18+/-0.12	0.77+/-0.06	0.67+/-0.05	0.69+/-0.04	0.59+/-0.02	1.17+/-0.14	0.80+/-0.05	0.77+/-0.03
Gram-pathogen feed	2.66+/-0.22	0.91+/-0.05	1.09+/-0.02	0.30+/-0.08	0.39+/-0.02	0.41+/-0.02	0.67+/-0.03	0.88+/-0.05	0.76+/-0.03	0.64+/-0.03
Fungus control	1.00+/-0.01	1.00+/-0.10	1.00+/-0.05	1.00+/-0.11	1.00+/-0.06	1.00+/-0.07	1.00+/-0.03	1.00+/-0.05	1.00+/-0.04	1.00+/-0.01
Fungus infected	0.87+/-0.15	0.78+/-0.03	0.85+/-0.05	1.10+/-0.30	1.07+/-0.06	1.30+/-0.06	1.43+/-0.06	0.92+/-0.19	0.97+/-0.09	2.31+/-0.06
Virus control	1.00+/-0.04	1.00+/-0.02	1.00+/-0.02	1.00+/-0.14	1.00+/-0.16	1.00+/-0.12	1.00+/-0.10	1.00+/-0.06	1.00+/-0.02	1.00+/-0.02
Virus infected	0.92+/-0.04	0.67+/-0.03	0.64+/-0.02	0.64+/-0.06	0.95+/-0.14	0.96+/-0.11	0.93+/-0.13	0.74+/-0.02	0.77+/-0.07	1.20+/-0.06
No stress control	1.00+/-0.02	1.00+/-0.05	1.00+/-0.02	1.00+/-0.06	1.00+/-0.11	1.00+/-0.04	1.00+/-0.05	1.00+/-0.08	1.00+/-0.07	1.00+/-0.05
Alarm pheromone	0.87+/-0.13	0.92+/-0.03	0.86+/-0.03	0.30+/-0.04	0.49+/-0.01	1.05+/-0.04	1.37+/-0.05	0.77+/-0.02	1.14+/-0.02	1.21+/-0.05
Post alarm pheromone	0.89+/-0.04	0.76+/-0.04	0.93+/-0.01	0.34+/-0.07	0.64+/-0.03	0.65+/-0.05	0.22+/-0.05	1.44+/-0.15	1.20+/-0.05	1.28+/-0.03
Heat stress	1.48+/-0.04	0.98+/-0.05	0.66+/-0.02	1.70+/-0.10	0.95+/-0.05	0.68+/-0.02	0.35+/-0.06	1.55+/-0.09	2.07+/-0.06	5.54+/-0.27
Post heat stress	0.78+/-0.13	0.61+/-0.02	0.67+/-0.04	0.31+/-0.07	0.85+/-0.04	0.74+/-0.06	0.64+/-0.05	0.92+/-0.02	1.01+/-0.07	0.26+/-0.01
Starvation	1.30+/-0.11	2.96+/-0.23	1.19+/-0.06	0.75+/-0.02	0.90+/-0.05	0.93+/-0.04	0.53+/-0.04	1.09+/-0.08	1.02+/-0.10	0.64+/-0.04

Table S6. Gut EST library statistics.

	TD2a library	TD2b library
Library ID	ID0AFF	ID0AAG
Tissue treatment	Control digestive tract	Gram – challenged digestive tract
		
Number of clones sequenced	5283	4043
Mean clone length	490 ± 160	602 ± 219
Median clone length	548	694
N seq. < 100 bp	210	197
Clones with no blastX hit (E≥10) [a]	737	633
% without hit	14,0%	15,7%
Clones with blastX hit (E<10)	4546 (4065*)	3410 (2921*)
Clones with non-sign. hits (10^{-3} – 10)	1408	936
Clones with sign. hits (E ≤ 10^{-3})	3138	2474
% with sign. hit	59,4%	61,2%
Uniprot hits from blastX (total)	2192	806
Uniprot hits from blastX (significant)	1486	578
N hits with more than 10 clones	56	50
N hits with more than 1% expression	6	12
N hits with more than 2% expression	1	9
Hits with GO annotation (%) [b]	79 %	80 %
Hits without GO annotation (total Fatigo)	312 (1495)	117 (579)
N Contigs (Ap v5 clustering)	2128	875
redundancy index (% contigs vs clones)	40.3 %	21.6 %
Nb specific contigs (control *vs* challenged)	1724	471
specificity rate (% specific vs total)	81%	54%
Fatigo analysis, total genes analysed	1495	579
Go class (4th level) [c]		
biopolymer metabolism (*0.005* – 0.088)	**21.0 %**	**11.3 %**
transport (*0.029* – 0.45)	**19.1 %**	**27.1 %**
establishment of localization (*0.038* – 0.64)	**19.2 %**	**27.0 %**
cellular metabolism	**78.4 %**	**71.7 %**
cell organization and biogenesis	**7.4 %**	**3.8 %**
primary metabolism	**67.7 %**	**62.3 %**
macromolecule metabolism	**46.9 %**	**44.0 %**
biosynthesis	**24.3 %**	**26.4 %**
catabolism	**14.3 %**	**12.6 %**

a : blastX performed through the blast parsing script runblastncbi (courtesy Laurent Duret), on Uniprot DB (May 2005, release UniprotKB 4.0, pbil server pbil.univ-lyon1.fr)

b : FATIGO links to GO ontology (June 2005, UniprotKB release 4.0, www.fatigo.org)

c : beneath % of class are given the *unadjusted* and adjusted (step-down min p) p values for library comparison, as computed by Fatigo (Fisher exact test). For non-discrimnant classes (no p given, p≥10%), only main classes are listed (>10% representation)

* : corrected for hit redundancy in blast results (low complexity or modular proteins)

Table S7. List of selected ESTs from the subtracted library. Results include the highest homologous match for each EST against the *A. pisum* NCBI refseq set (blastX), the publicly available *A. Pisum* ESTs (blastN), and the SwissProt databases (blastX). ESTs have been selected according to either apparent differential gene expression or to their similarities to putative immune-related genes. Differential expression status of ESTs was analyzed using a dot-blot experiment as described in the materials and methods. The change in gene expression after *E. coli* infection relative to the untreated aphid is given as qualitative data (+: overexpression; -: underexpression; =: no differential gene expression). Descriptions for aphid matches are based on *A. pisum* EST cluster descriptions, available at the *Acyrthosiphon Pisum* EST Database [11], and/or on refseq descriptions.

EST			Aphid Hits				Swiss Prot Hits		
GenBank Acc. Num.	Length	Relative expression	*A. pisum* gene (evalue)	*A. pisum* EST (evalue)	*A. pisum* EST cluster	Description based on gene or EST	Swiss Prot ID (evalue)	Organism	Description
GD186025	237	-	no hit	EX619631 (e-104)	APG20589		no hit		
GD186052	629	-	ACYPI006141 (5e-31)	FF333609 (0.0)	APG03947	similar to eukaryotic translation initiation factor 5	Q9VXK6 (9e-26)	*Drosophila melanogaster*	eukaryotic translation initiation factor 5
GD186090	299	-	ACYPI004024 (3e-11)	FF338507 (e-106)	APG01676	transport, small GTPase mediated signal transduction	Q9VZ23 (2e-07)	*Drosophila melanogaster*	GTP-binding nuclear protein Ran
GD186102	425	-	no hit	DY224167	APD07979		no hit		
GD185918	445	+	no hit	FF317856	APG05151		no hit		
GD185990	539	+	ACYPI005949 (5e-99)	FF334362 (0.0)	APG08004	Rps3 ribosomal protein S3	P48153 (2e-81)	*Manduca sexta*	40S ribosomal protein S3
GD186053	320	+	ACYPI006749 (8e-34)	EX620733 (e-169)	APG08261	methionine biosynthesis	no hit		
GD186057	748	+	ACYPI006004 (2e-189)	FF324610 (0.0)	APG11294	similar to latent nuclear antigen	Q8IMP6 (7e-23)	*Drosophila melanogaster*	protein SPT2 homolog
GD186146	540	+	ACYPI003334 (6e-62)	EX607065 (0.0)	APG09767		Q6GMF8 (5e-46)	*Danio rerio*	rhomboid family member 1
GD186208	397	+	no hit	FF317879 (0.0)	APG01972	nuclear mRNA splicing	no hit		
GD186213	399	+	no hit	EX653226 (e-113)	APG17996		no hit		
GD186220	314	+	ACYPI000343 (8e-59)	CN762809 (e-155)	APD04091	similar to ubiquitin specific protease 7	Q9VYQ8 (5e-17)	*Drosophila melanogaster*	ubiquitin carboxyl-terminal hydrolase 7
GD186223	571	+	ACYPI007035 (4e-85)	FF317872 (0.0)	APG06079	similar to coiled-coil domain-containing protein 132	Q8CI71 (1e-19)	*Mus musculus*	coiled-coil domain-containing protein 132
GD186225	333	+	no hit	no hit			no hit		
GD186231	379	+	ACYPI003578 (2e-58)	FF314239 (0.0)	APG03794	similar to signal recognition particle 72 kDa protein	P33731 (7e-15)	*Canis lupus familiaris*	signal recognition particle 72 kDa protein
GD186337	674	+	no hit	FF333448 (e-115)	APG13156	amino acid transport	Q54S12 (0.12)	*Dictyostelium discoideum*	transmembrane protein 104 homolog
GD186362	397	+	no hit	DY223529 (e-103)	APD03776		no hit		
GD186383	375	+	no hit	FF305066 (0.0)	APG03552	protein amino acid phosphorylation	Q16513 (0.003)	*Homo sapiens*	serine/threonine-protein kinase N2
GD186386	361	+	no hit	FF330011	APG09318		no hit		

EST			Aphid Hits				Swiss Prot Hits		
GenBank Acc. Num.	Length	Relative expression	*A. pisum* gene (evalue)	*A. pisum* EST (evalue)	*A. pisum* EST cluster	Description based on gene or EST	Swiss Prot ID (evalue)	Organism	Description
				(e-115)					
GD185976	283	=	ACYPI002027 (2e-58)	CV838512 (e-154)	APD02867	thrombospondin-like	Q9C0I4 (1e-16)	*Homo sapiens*	thrombospondin type-1 domain-containing protein 7B
GD186047	347	=	ACYPI000111 (1e-68)	no hit		similar to adam	Q9VAC5 (2e-53)	*Drosophila melanogaster*	Adam 17-like protease
GD186063	1130	=	ACYPI000719 (3e-136)	FF336689 (0.0)	APG09006		P13284 (7e-13)	*Homo sapiens*	gamma-interferon-inducible lysosomal thiol reductase precursor
GD186100	284	=	ACYPI008118 (4e-102)	FF315088 (0.0)	APG16211	similar to tyrosine protein kinase	Q24592 (2e-08)	*Drosophila melanogaster*	tyrosine-protein kinase hopscotch (Jak)
GD186119	325	=	ACYPI005016 (2e-51)	FF320908 (e-176)	APG08844	serine protease inhibitor (serpin 1)	P48594 (2e-17)	*Homo sapiens*	serpin B4
GD186120	353	=	ACYPI005540 (6e-63)	CN759023 (0.0)	APD04504	MAPKKK cascade	Q5E9X2 (3e-45)	*Bos taurus*	dual specificity mitogen-activated protein kinase 6
GD186131	517	=	ACYPI005016 (5e-86)	FF336182 (0.0)	APG07692	serine-protease inhibitor (serpin 1)	P48594 (2e-17)	*Homo sapiens*	serpin B4
GD186142	328	=	no hit	EX639251 (e-110)	APG19100	receptor activity	Q5XIN3 (1e-07)	*Rattus norvegicus*	TRAF3-interacting protein 1
GD186162	330	=	ACYPI003236 (3e-59)	no hit		similar to sunday driver	Q9ESN9 (9e-34)	*Mus musculus*	JNK-interacting protein 3
GD186173	245	=	no hit	CN761251 (e-119)	APD04293	programs cell death	Q3UU35 (0.8)	*Mus musculus*	Ovostatin homolog
GD186257	345	=	ACYPI006651 (7e-32)	CN764436 (e-163)	APD05533	similar to membrane-associated LPS-inducible TNF alpha factor	Q8QGW7 (7e-10)	*Gallus gallus*	LPS-induced TNF-alpha factor homolog
GD186266	346	=	ACYPI002465 (3e-28)	CN756373 (e-173)	APD05909	similar to macrophage migration inhibitory factor	P91850 (3e-08)	*Brugia malayi*	Macrophage migration inhibitory factor homolog
GD186363	268	=	no hit	EX650498 (e-126)	APG05258	sugar binding	no hit		

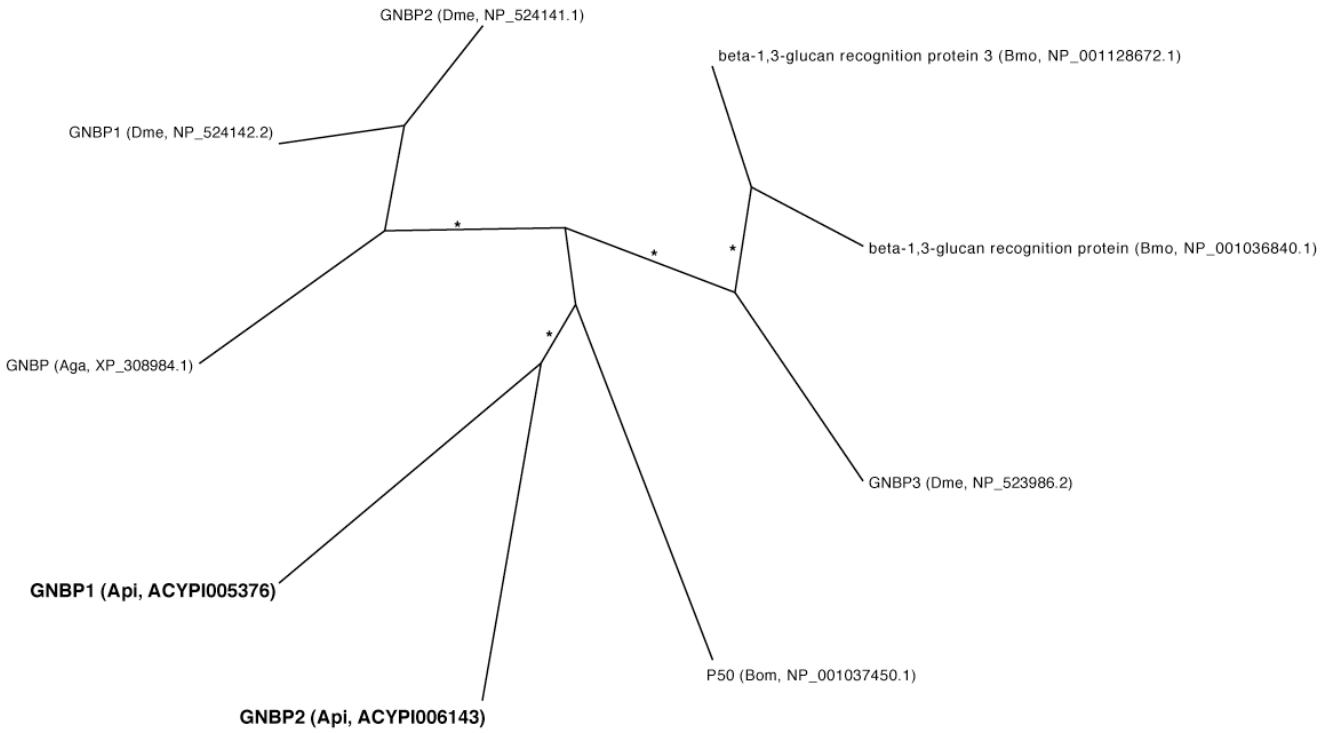
Table S7. List of selected ESTs from the subtracted library. Results include the highest homologous match for each EST against the *A. pisum* NCBI refseq set (blastX), the publicly available *A. Pisum* ESTs (blastN), and the SwissProt databases (blastX). ESTs have been selected according to either apparent differential gene expression or to their similarities to putative immune-related genes. Differential expression status of ESTs was analyzed using a dot-blot experiment as described in the materials and methods. The change in gene expression after *E. coli* infection relative to the untreated aphid is given as qualitative data (+: overexpression; -: underexpression; =: no differential gene expression). Descriptions for aphid matches are based on *A. pisum* EST cluster descriptions, available at the *Acyrthosiphon Pisum* EST Database [11], and/or on refseq descriptions.

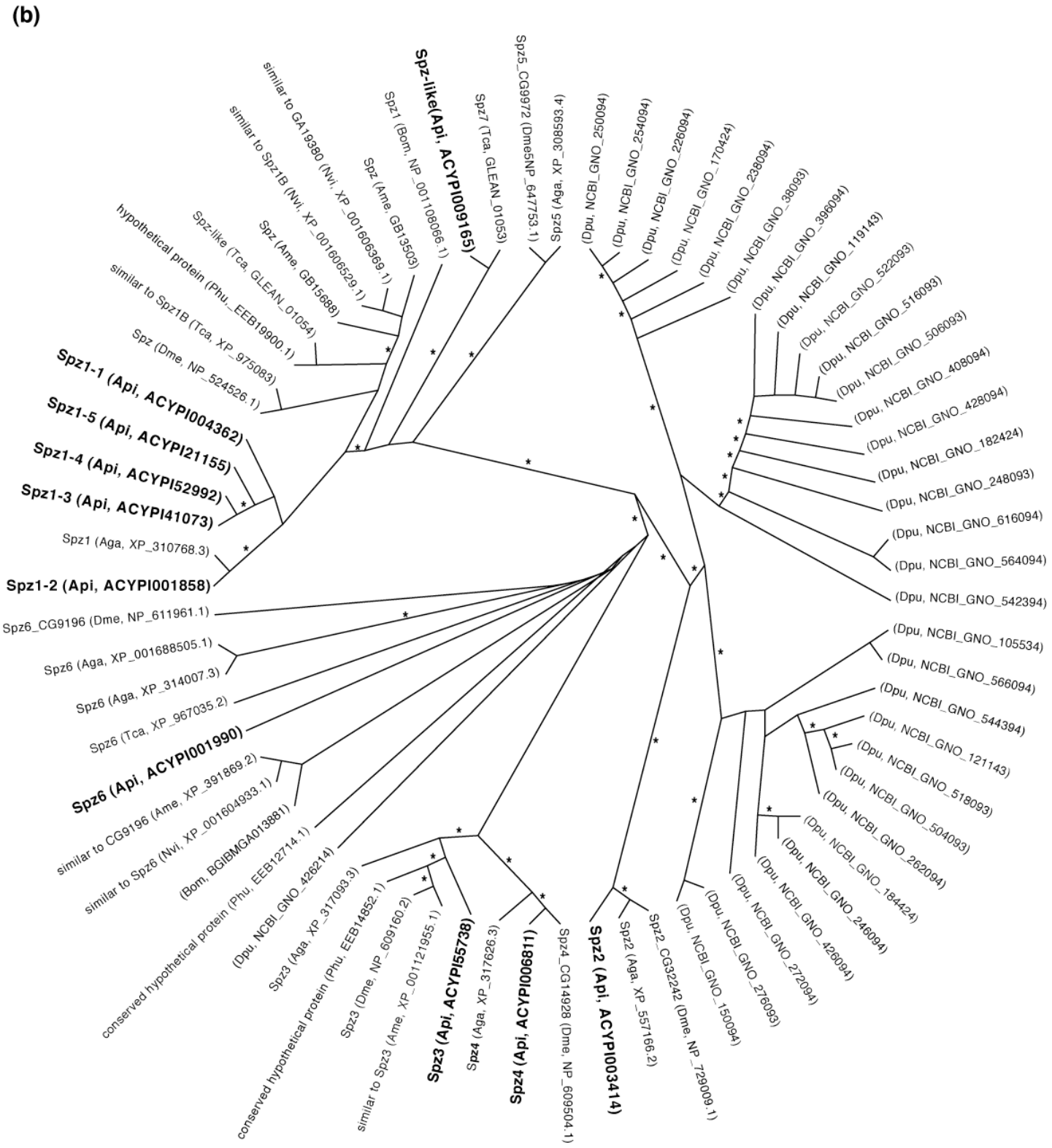
EST			Aphid Hits				Swiss Prot Hits		
GenBank Acc. Num.	Length	Relative expression	*A. pisum* gene (evalue)	*A. pisum* EST (evalue)	*A. pisum* EST cluster	Description based on gene or EST	Swiss Prot ID (evalue)	Organism	Description
GD186025	237	-	no hit	EX619631 (e-104)	APG20589		no hit		
GD186052	629	-	ACYPI006141 (5e-31)	FF333609 (0.0)	APG03947	similar to eukaryotic translation initiation factor 5	Q9VXK6 (9e-26)	*Drosophila melanogaster*	eukaryotic translation initiation factor 5
GD186090	299	-	ACYPI004024 (3e-11)	FF338507 (e-106)	APG01676	transport, small GTPase mediated signal transduction	Q9VZ23 (2e-07)	*Drosophila melanogaster*	GTP-binding nuclear protein Ran
GD186102	425	-	no hit	DY224167	APD07979		no hit		
GD185918	445	+	no hit	FF317856	APG05151		no hit		
GD185990	539	+	ACYPI005949 (5e-99)	FF334362 (0.0)	APG08004	Rps3 ribosomal protein S3	P48153 (2e-81)	*Manduca sexta*	40S ribosomal protein S3
GD186053	320	+	ACYPI006749 (8e-34)	EX620733 (e-169)	APG08261	methionine biosynthesis	no hit		
GD186057	748	+	ACYPI006004 (2e-189)	FF324610 (0.0)	APG11294	similar to latent nuclear antigen	Q8IMP6 (7e-23)	*Drosophila melanogaster*	protein SPT2 homolog
GD186146	540	+	ACYPI003334 (6e-62)	EX607065 (0.0)	APG09767		Q6GMF8 (5e-46)	*Danio rerio*	rhomboid family member 1
GD186208	397	+	no hit	FF317879 (0.0)	APG01972	nuclear mRNA splicing	no hit		
GD186213	399	+	no hit	EX653226 (e-113)	APG17996		no hit		
GD186220	314	+	ACYPI000343 (8e-59)	CN762809 (e-155)	APD04091	similar to ubiquitin specific protease 7	Q9VYQ8 (5e-17)	*Drosophila melanogaster*	ubiquitin carboxyl-terminal hydrolase 7
GD186223	571	+	ACYPI007035 (4e-85)	FF317872 (0.0)	APG06079	similar to coiled-coil domain-containing protein 132	Q8CI71 (1e-19)	*Mus musculus*	coiled-coil domain-containing protein 132
GD186225	333	+	no hit	no hit			no hit		
GD186231	379	+	ACYPI003578 (2e-58)	FF314239 (0.0)	APG03794	similar to signal recognition particle 72 kDa protein	P33731 (7e-15)	*Canis lupus familiaris*	signal recognition particle 72 kDa protein
GD186337	674	+	no hit	FF333448 (e-115)	APG13156	amino acid transport	Q54S12 (0.12)	*Dictyostelium discoideum*	transmembrane protein 104 homolog
GD186362	397	+	no hit	DY223529 (e-103)	APD03776		no hit		
GD186383	375	+	no hit	FF305066 (0.0)	APG03552	protein amino acid phosphorylation	Q16513 (0.003)	*Homo sapiens*	serine/threonine-protein kinase N2
GD186386	361	+	no hit	FF330011	APG09318		no hit		

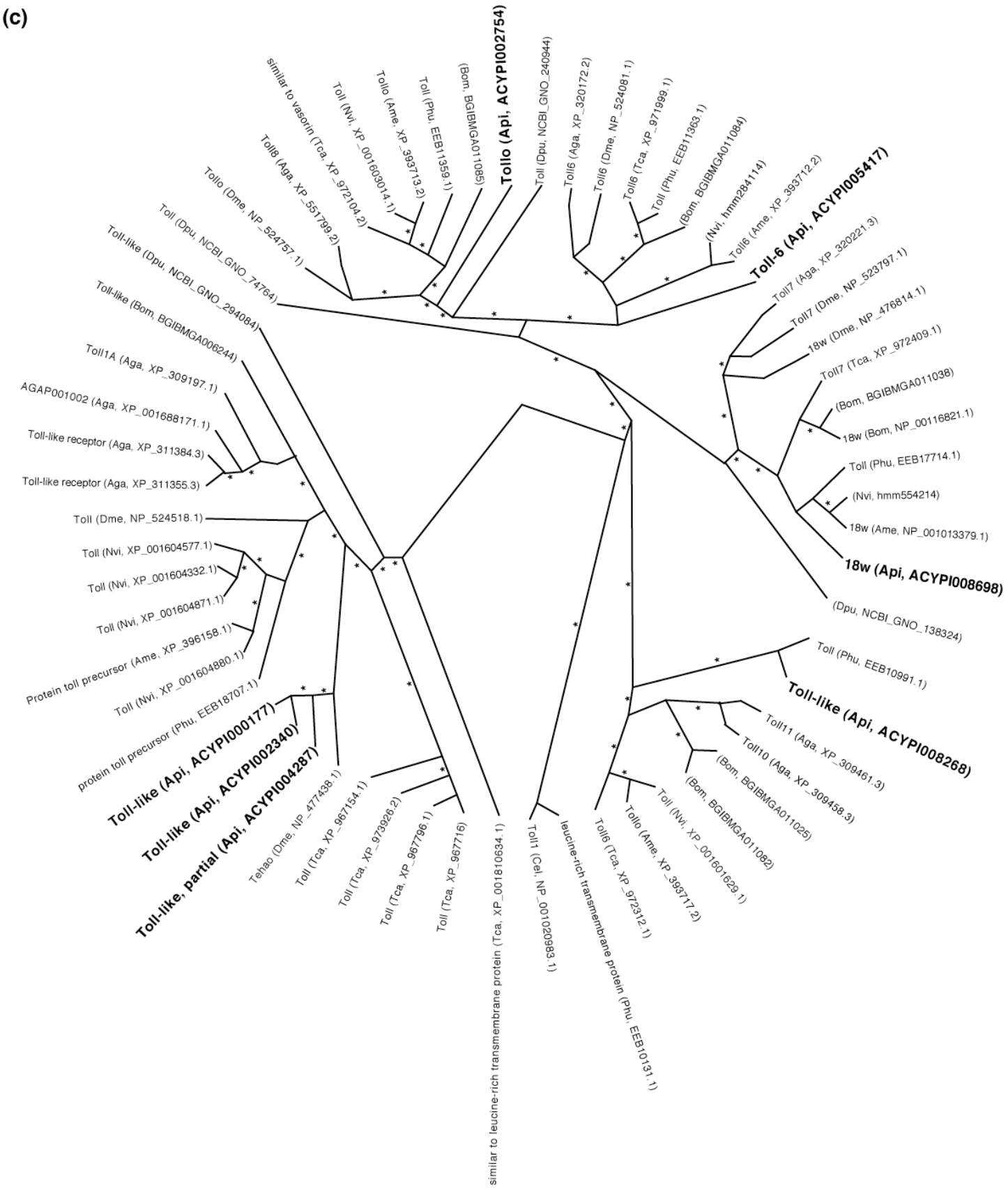
EST			Aphid Hits				Swiss Prot Hits		
GenBank Acc. Num.	Length	Relative expression	*A. pisum* gene (evalue)	*A. pisum* EST (evalue)	*A. pisum* EST cluster	Description based on gene or EST	Swiss Prot ID (evalue)	Organism	Description
				(e-115)					
GD185976	283	=	ACYPI002027 (2e-58)	CV838512 (e-154)	APD02867	thrombospondin-like	Q9C0I4 (1e-16)	*Homo sapiens*	thrombospondin type-1 domain-containing protein 7B
GD186047	347	=	ACYPI000111 (1e-68)	no hit		similar to adam	Q9VAC5 (2e-53)	*Drosophila melanogaster*	Adam 17-like protease
GD186063	1130	=	ACYPI000719 (3e-136)	FF336689 (0.0)	APG09006		P13284 (7e-13)	*Homo sapiens*	gamma-interferon-inducible lysosomal thiol reductase precursor
GD186100	284	=	ACYPI008118 (4e-102)	FF315088 (0.0)	APG16211	similar to tyrosine protein kinase	Q24592 (2e-08)	*Drosophila melanogaster*	tyrosine-protein kinase hopscotch (Jak)
GD186119	325	=	ACYPI005016 (2e-51)	FF320908 (e-176)	APG08844	serine protease inhibitor (serpin 1)	P48594 (2e-17)	*Homo sapiens*	serpin B4
GD186120	353	=	ACYPI005540 (6e-63)	CN759023 (0.0)	APD04504	MAPKKK cascade	Q5E9X2 (3e-45)	*Bos taurus*	dual specificity mitogen-activated protein kinase 6
GD186131	517	=	ACYPI005016 (5e-86)	FF336182 (0.0)	APG07692	serine-protease inhibitor (serpin 1)	P48594 (2e-17)	*Homo sapiens*	serpin B4
GD186142	328	=	no hit	EX639251 (e-110)	APG19100	receptor activity	Q5XIN3 (1e-07)	*Rattus norvegicus*	TRAF3-interacting protein 1
GD186162	330	=	ACYPI003236 (3e-59)	no hit		similar to sunday driver	Q9ESN9 (9e-34)	*Mus musculus*	JNK-interacting protein 3
GD186173	245	=	no hit	CN761251 (e-119)	APD04293	programs cell death	Q3UU35 (0.8)	*Mus musculus*	Ovostatin homolog
GD186257	345	=	ACYPI006651 (7e-32)	CN764436 (e-163)	APD05533	similar to membrane-associated LPS-inducible TNF alpha factor	Q8QGW7 (7e-10)	*Gallus gallus*	LPS-induced TNF-alpha factor homolog
GD186266	346	=	ACYPI002465 (3e-28)	CN756373 (e-173)	APD05909	similar to macrophage migration inhibitory factor	P91850 (3e-08)	*Brugia malayi*	Macrophage migration inhibitory factor homolog
GD186363	268	=	no hit	EX650498 (e-126)	APG05258	sugar binding	no hit		

Figure S1. Maximum likelihood phylogenies of selected immune gene families. (a) Gram-negative binding proteins; (b) Spätzles; (c) Tolls, note that there is no support for *A. pisum* tolls clading with *D. melanogaster* Tehao; (d) Lysozymes; (e) Teps; (f) Prophenoloxidases; and, (g) low molecular weight heat shock proteins (Hsp83, Hsp90 not included). * represents approximate likelihood ratio test (aLRT) support > 80. (Aga: *Anopheles gambiae*; Ame: *Apis Mellifera*; Api: *Acythrosiphon pisum*; Bom: *Bombyx mori*; Cel: *Caenorhabditis elegans*; Cin: *Ciona_intestinalis*; Dme = *Drosophila melanogaster*: Dpu = *Daphnia pulex*; Nvi: *Nasonia vitripennis*; Phu: *Pediculus humanus*; Tca: *Tribolium castaneum*.)

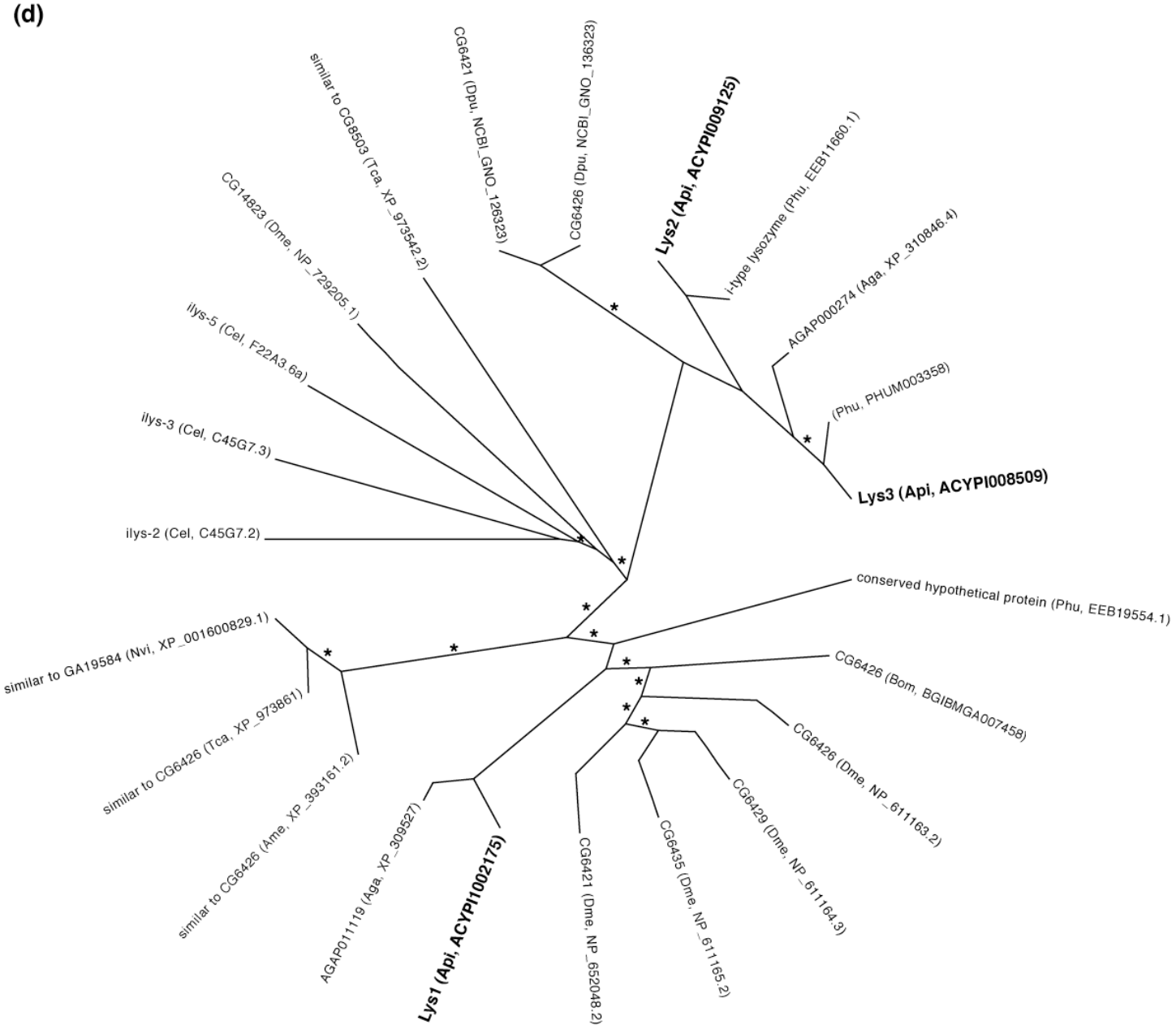
(a)



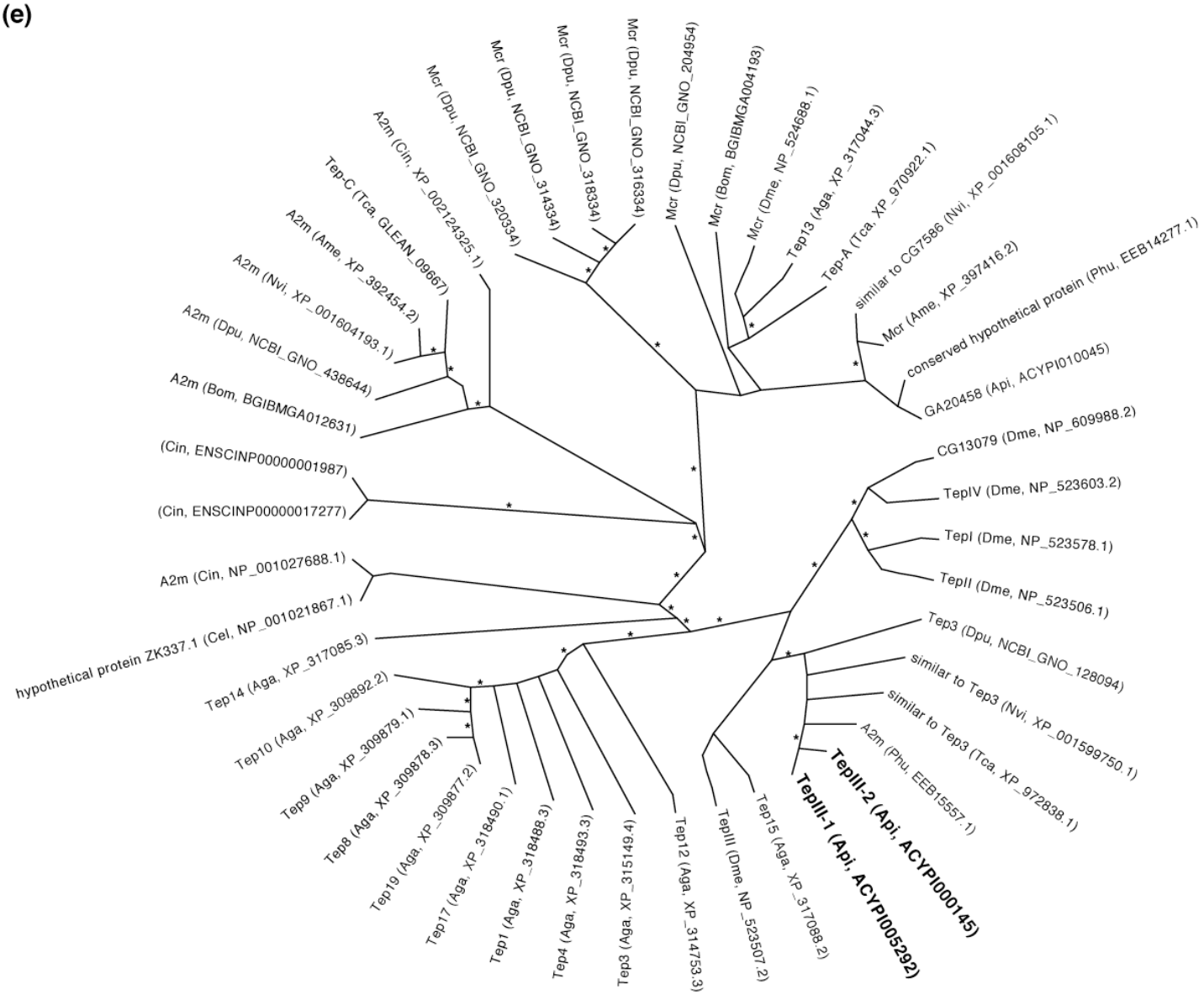

(b)
Spz-like(Api, ACYPI009165)
Spz5_CG9972 (Dme5NP_647753.1)
Spz25 (Aga, XP_308593.4)
Spz7 (Tca, GLEAN_01053)
Spz1 (Bom, NP_001108066.1)
similar to GA19380 (Nvi, XP_001606369.1)
Spz (Ame, GB13509)
similar to Spz1B (Nvi, XP_001606529.1)
Spz (Ame, GB15688)
Spz-like (Tca, GLEAN_01054)
hypothetical protein (Phu, EEB19900.1)
similar to Spz1B (Tca, XP_975083)
Spz (Dme, NP_524526.1)
Spz1-1 (Api, ACYPI004362)
Spz1-5 (Api, ACYPI21155)
Spz1-4 (Api, ACYPI52992)
Spz1-3 (Api, ACYPI41073)
Spz1 (Aga, XP_310768.3)
Spz1-2 (Api, ACYPI001858)
Spz6_CG9196 (Dme, NP_611961.1)
Spz6 (Aga, XP_001688505.1)
Spz6 (Aga, XP_314007.3)
Spz6 (Tca, XP_967035.2)
Spz6 (Api, ACYPI001990)
similar to CG9196 (Ame, XP_391869.2)
similar to Spz6 (Nvi, XP_001604933.1)
(Bom, BGIBMGA013881)
conserved hypothetical protein (Phu, EEB12714.1)
(Dpu, NCBI_GNO_426214)
Spz3 (Aga, XP_317093.3)
conserved hypothetical protein (Phu, EEB14852.1)
Spz3 (Dme, NP_609160.2)
similar to Spz3 (Ame, XP_001121955.1)
Spz3 (Api, ACYPI55738)
Spz4 (Aga, XP_317626.3)
Spz4 (Api, ACYPI006811)
Spz4_CG14928 (Dme, NP_609504.1)
Spz2 (Api, ACYPI003414)
Spz2 (Aga, XP_557166.2)
Spz2_CG32242 (Dme, NP_729009.1)
(Dpu, NCBI_GNO_150094)
(Dpu, NCBI_GNO_276093)
(Dpu, NCBI_GNO_272094)
(Dpu, NCBI_GNO_426094)
(Dpu, NCBI_GNO_246094)
(Dpu, NCBI_GNO_184424)
(Dpu, NCBI_GNO_262094)
(Dpu, NCBI_GNO_504093)
(Dpu, NCBI_GNO_518093)
(Dpu, NCBI_GNO_121143)
(Dpu, NCBI_GNO_544394)
(Dpu, NCBI_GNO_566094)
(Dpu, NCBI_GNO_105534)
(Dpu, NCBI_GNO_542394)
(Dpu, NCBI_GNO_564094)
(Dpu, NCBI_GNO_616094)
(Dpu, NCBI_GNO_248093)
(Dpu, NCBI_GNO_182424)
(Dpu, NCBI_GNO_428094)
(Dpu, NCBI_GNO_408094)
(Dpu, NCBI_GNO_506093)
(Dpu, NCBI_GNO_516093)
(Dpu, NCBI_GNO_522093)
(Dpu, NCBI_GNO_119143)
(Dpu, NCBI_GNO_396094)
(Dpu, NCBI_GNO_38093)
(Dpu, NCBI_GNO_238094)
(Dpu, NCBI_GNO_170424)
(Dpu, NCBI_GNO_226094)
(Dpu, NCBI_GNO_254094)
(Dpu, NCBI_GNO_250094)


(c)
Tollo (Api, ACYPI002754)
Toll (Dpu, NCBI_GNO_240944)
Toll6 (Aga, XP_320172.2)
Toll6 (Dme, NP_524081.1)
Toll6 (Tca, XP_971999.1)
Toll (Phu, EEB11363.1)
(Bom, BGIBMGA011084)
(Nvi, hmm284114)
Toll6 (Ame, XP_393712.2)
Toll-6 (Api, ACYPI005417)
Toll7 (Aga, XP_320221.3)
Toll7 (Dme, NP_523797.1)
18w (Dme, NP_476814.1)
Toll7 (Tca, XP_972409.1)
(Bom, BGIBMGA011038)
18w (Bom, NP_00116821.1)
Toll (Phu, EEB17714.1)
(Nvi, hmm554214)
18w (Ame, NP_001013379.1)
18w (Api, ACYPI008698)
(Dpu, NCBI_GNO_138324)
Toll (Phu, EEB10991.1)
Toll-like (Api, ACYPI008268)
Toll11 (Aga, XP_309461.3)
Toll10 (Aga, XP_309458.3)
(Bom, BGIBMGA011025)
(Bom, BGIBMGA011082)
Toll (Nvi, XP_001601629.1)
Tollo (Ame, XP_393717.2)
Toll6 (Tca, XP_972312.1)
leucine-rich transmembrane protein (Phu, EEB10131.1)
Toll1 (Cel, NP_001020983.1)
similar to leucine-rich transmembrane protein (Tca, XP_001810634.1)
Toll (Tca, XP_967716)
Toll (Tca, XP_967796.1)
Toll (Tca, XP_973926.2)
Toll (Tca, XP_967154.1)
Tehao (Dme, NP_477438.1)
Toll-like, partial (Api, ACYPI004287)
Toll-like (Api, ACYPI002340)
Toll-like (Api, ACYPI000177)
protein toll precursor (Phu, EEB18707.1)
Toll (Nvi, XP_001604880.1)
Protein toll precursor (Ame, XP_396158.1)
Toll (Nvi, XP_001604871.1)
Toll (Nvi, XP_001604332.1)
Toll (Nvi, XP_001604577.1)
Toll (Dme, NP_524518.1)
Toll-like receptor (Aga, XP_311355.3)
Toll-like receptor (Aga, XP_311384.3)
AGAP001002 (Aga, XP_001688171.1)
Toll1A (Aga, XP_309197.1)
Toll-like (Bom, BGIBMGA006244)
Toll-like (Dpu, NCBI_GNO_294084)
Toll (Dpu, NCBI_GNO_74764)
Tollo (Dme, NP_524757.1)
Toll8 (Aga, XP_551799.2)
similar to vasorin (Tca, XP_972104.2)
Toll (Nvi, XP_001603014.1)
Tollo (Ame, XP_393713.2)
Toll (Phu, EEB11359.1)
(Bom, BGIBMGA011085)

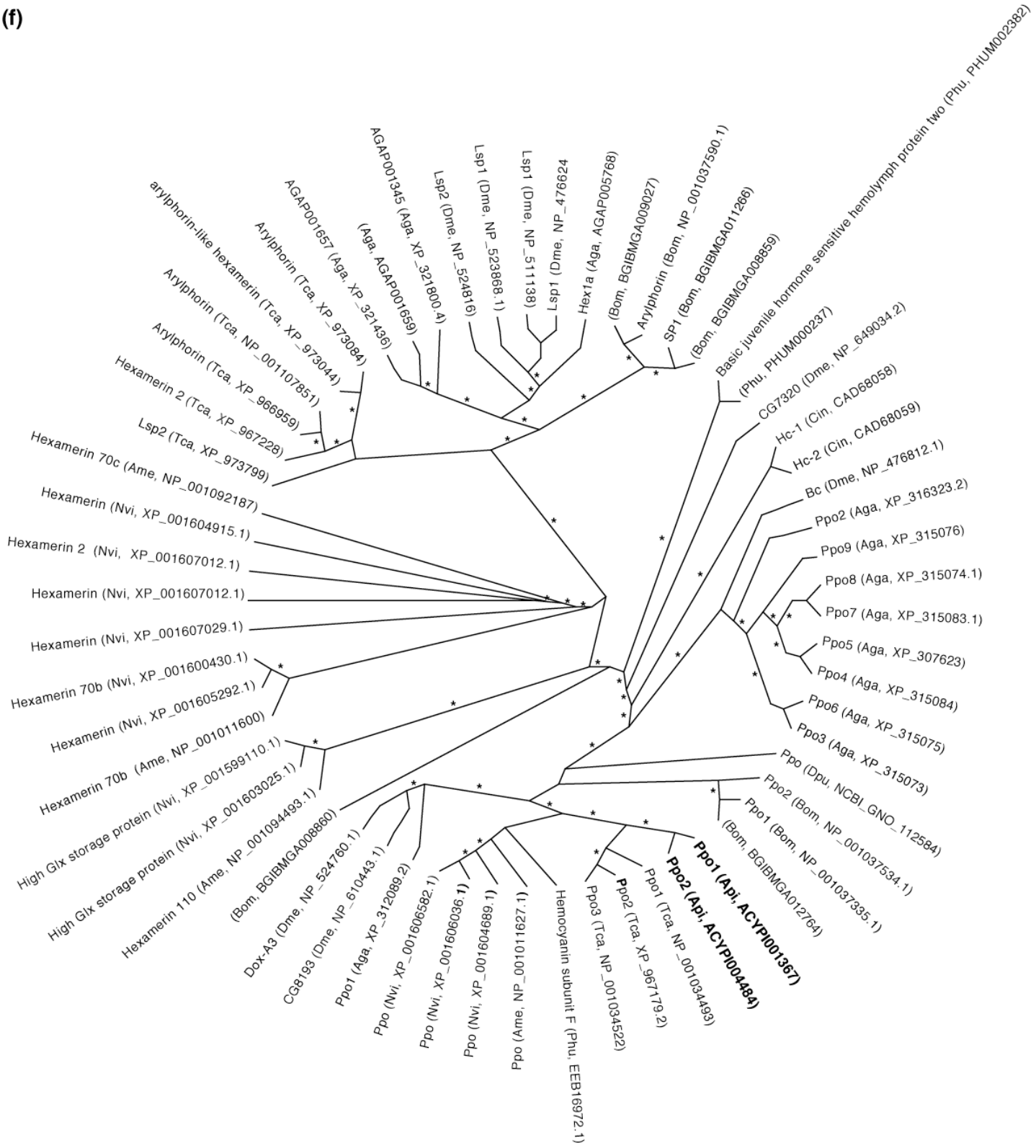
(d)


similar to CG8503 (Tca, XP_973542.2)
CG6421 (Dpu, NCBI_GNO_126323)
CG6426 (Dpu, NCBI_GNO_136323)
Lys2 (Api, ACYPI009125)
i-type lysozyme (Phu, EEB11660.1)
CG14823 (Dme, NP_729205.1)
AGAP000274 (Aga, XP_310846.4)
ilys-5 (Cel, F22A3.6a)
(Phu, PHUM003358)
ilys-3 (Cel, C45G7.3)
Lys3 (Api, ACYPI008509)
ilys-2 (Cel, C45G7.2)
conserved hypothetical protein (Phu, EEB19554.1)
similar to GA19584 (Nvi, XP_001600829.1)
CG6426 (Bom, BGIBMGA007458)
similar to CG6426 (Tca, XP_973861)
CG6426 (Dme, NP_611163.2)
similar to CG6426 (Ame, XP_393161.2)
CG6429 (Dme, NP_611164.3)
AGAP011119 (Aga, XP_309527)
CG6435 (Dme, NP_611165.2)
Lys1 (Api, ACYPI002175)
CG6421 (Dme, NP_652048.2)

(e)


Mcr (Dpu, NCBI_GNO_316334)
Mcr (Dpu, NCBI_GNO_318334)
Mcr (Dpu, NCBI_GNO_314334)
Mcr (Dpu, NCBI_GNO_320334)
A2m (Cin, XP_002124325.1)
Tep-C (Tca, GLEAN_09867)
A2m (Ame, XP_392454.2)
A2m (Nvi, XP_001604193.1)
A2m (Dpu, NCBI_GNO_438644)
A2m (Bom, BGIBMGA012631)
(Cin, ENSCINP00000001987)
(Cin, ENSCINP00000017277)
A2m (Cin, NP_001027688.1)
hypothetical protein ZK337.1 (Cel, NP_001021867.1)
Tep14 (Aga, XP_317085.3)
Tep10 (Aga, XP_309892.2)
Tep9 (Aga, XP_309879.1)
Tep8 (Aga, XP_309878.3)
Tep19 (Aga, XP_309877.2)
Tep17 (Aga, XP_318490.1)
Tep1 (Aga, XP_318488.3)
Tep4 (Aga, XP_318493.3)
Tep3 (Aga, XP_315149.4)
Tep12 (Aga, XP_314753.3)
TepIII (Dme, NP_523507.2)
Tep15 (Aga, XP_317088.2)
TepIII-1 (Api, ACYPI005292)
TepIII-2 (Api, ACYPI000145)
A2m (Phu, EEB15557.1)
similar to Tep3 (Tca, XP_972838.1)
similar to Tep3 (Nvi, XP_001599750.1)
Tep3 (Dpu, NCBI_GNO_128094)
TepII (Dme, NP_523506.1)
TepI (Dme, NP_523578.1)
TepIV (Dme, NP_523603.2)
CG13079 (Dme, NP_609988.2)
GA20458 (Api, ACYPI010045)
conserved hypothetical protein (Phu, EEB14277.1)
Mcr (Ame, XP_397416.2)
similar to CG7586 (Nvi, XP_001606105.1)
Tep-A (Tca, XP_970922.1)
Tep13 (Aga, XP_317044.3)
Mcr (Dme, NP_524688.1)
Mcr (Bom, BGIBMGA004193)
Mcr (Dpu, NCBI_GNO_204954)

(f)


Basic juvenile hormone sensitive hemolymph protein two (Phu, PHUM002382)
(Phu, PHUM000237)
CG7320 (Dme, NP_649034.2)
Hc-1 (Cin, CAD68058)
Hc-2 (Cin, CAD68059)
Bc (Dme, NP_476812.1)
Ppo2 (Aga, XP_316323.2)
Ppo9 (Aga, XP_315076)
Ppo8 (Aga, XP_315074.1)
Ppo7 (Aga, XP_315083.1)
Ppo5 (Aga, XP_307623)
Ppo4 (Aga, XP_315084)
Ppo6 (Aga, XP_315075)
Ppo3 (Aga, XP_315073)
Ppo (Dpu, NCBI_GNO_112584)
Ppo2 (Bom, NP_001037534.1)
Ppo1 (Bom, NP_001037335.1)
(Bom, BGIBMGA012764)
Ppo1 (Api, ACYPI001367)
Ppo2 (Api, ACYPI004484)
Ppo1 (Tca, NP_001034493)
Ppo2 (Tca, XP_967179.2)
Ppo3 (Tca, NP_001034522)
Hemocyanin subunit F (Phu, EEB16972.1)
Ppo (Ame, NP_001011627.1)
Ppo (Nvi, XP_001604689.1)
Ppo (Nvi, XP_001606036.1)
Ppo (Nvi, XP_001606582.1)
Ppo1 (Aga, XP_312089.2)
CG8193 (Dme, NP_610443.1)
Dox-A3 (Dme, NP_524760.1)
(Bom, BGIBMGA008860)
Hexamerin 110 (Ame, NP_001094493.1)
High Glx storage protein (Nvi, XP_001603025.1)
High Glx storage protein (Nvi, XP_001599110.1)
Hexamerin 70b (Ame, NP_001011600)
Hexamerin (Nvi, XP_001605292.1)
Hexamerin 70b (Nvi, XP_001600430.1)
Hexamerin (Nvi, XP_001607029.1)
Hexamerin (Nvi, XP_001607012.1)
Hexamerin 2 (Nvi, XP_001607012.1)
Hexamerin (Nvi, XP_001604915.1)
Hexamerin 70c (Ame, NP_001092187)
Lsp2 (Tca, XP_973799)
Hexamerin 2 (Tca, XP_967228)
Arylphorin (Tca, XP_966959)
Arylphorin (Tca, NP_001107851)
arylphorin-like hexamerin (Tca, XP_973044)
Arylphorin (Tca, XP_973084)
AGAP001657 (Aga, XP_321436)
(Aga, AGAP001659)
AGAP001345 (Aga, XP_321800.4)
Lsp2 (Dme, NP_524816)
Lsp1 (Dme, NP_523868.1)
Lsp1 (Dme, NP_511138)
Lsp1 (Dme, NP_476624
Hex1a (Aga, AGAP005768)
(Bom, BGIBMGA009027)
Arylphorin (Bom, NP_001037590.1)
SP1 (Bom, BGIBMGA011266)
(Bom, BGIBMGA008859)

(g)


Hsp70Ba (Api, ACYPI008763)
Hsp70Ab (Api, ACYPI009961)
Hsp70Aa (Api, ACYPI009171)
Hsc70 (Api, ACYPI007166)
Hsc5 (Api, ACYPI004693)
Hsc70-1 (Api, ACYPI000474)
Hsc70-2 (Api, ACYPI004809)
Hsc70Cb (Api, ACYPI004544)
Hsp14 (Api, ACYPI002719)
Hsp60 (Api, ACYPI009253)
Hsp21.4 (Api, ACYPI003907)
hypothetical protein F44e5.4 (Cel, NP_496509.1)
Hsp70 (Cel, NP_492485.1)
Hsp70 (Cin, NP_001029006.1)
(Dpu, NCBI_GNO_546344)
Hsc71 (Phu, EEB17736.1)
(Bom, BGIBMGA001635)
similar to Hsc70 (Cin, XP_002127103.1)
(Cin, ENSCINP00000019498)
(Dpu, NCBI_GNO_762034)
similar to Hsc3 (Ame, XP_393090.3)
similar to ENSANGP00000012893 (Nvi, XP_001606463.1)
Hsp70 (Bom, NP_001036837.1)
AGAP004192 (Aga, XP_313085.3)
(Tca, GLEAN_04425)
Hsc70 (Phu, EEB17838.1)
Hsp3 isoA (Dme, NP_727563.1)
Hsp3 (Cel, NP_509019.1)
Hsp4 (Cel, NP_495536.1)
similar to grp75 (Cin, XP_002121036.1)
Hsp70 (Dpu, NCBI_GNO_618084)
AGAP010876 (Aga, XP_309825.4)
Hsc5 (Dme, EEB15346.1)
Hsc70 (Dme, NP_523741.2)
similar to conserved hypothetical protein (Tca, XP_975386.1)
similar to CG8542 (Nvi, XP_001599525.1)
similar to Hsc5 (Ame, XP_392147.2)
Hsp6 (Cel, NP_504291.1)
(Dpu, NCBI_GNO_106854)
hypothetical protein F11F1.1 (Cel, NP_001021249.1)
similar to Hsc1 (Tca, XP_968075.2)
similar to Hsp70 (Tca, XP_971162.2)
similar to predicted protein (Tca, XP_975210.1)
AGAP007682 (Aga, XP_001237047.2)
Stch (Cel, NP_495808.2)
similar to Hsp (Nvi, XP_001599254.1)
(Dpu, NCBI_GNO_102434)
150 kDa oxygen-regulated protein precursor (Phu, EEB20143.1)
(Bom, BGIBMGA008551)
(Cin, ENSCINT00000007671)
similar to ENSANGP00000015293 (Nvi, XP_001607146.1)
similar to Hsc70Cb (Ame, XP_623199.1)
similar to AGAP010331 (Tca, XP_975279.2)
Hsp 105 kDa (Phu, EEB18266.1)
(Bom, BGIBMGA001218)
Hsc70Cb isoA (Dmel, NP_648687.1)
AGAP010331 (Aga, XP_315285.3)
Hsp70 (Dpu, NCBI_GNO_370774)
hypothetical protein C30C11.4 (Cel, NP_498868.1)
Hsp SSB (Phu, EEB14158.1)
Hsp70 (Dpu, NCBI_GNO_322413)
(Bom, BGIBMGA008574)
similar to Hsp1 (Tca, XP_968760.1)
AGAP007162 (Aga, XP_308606.2)
AGAP007161 (Aga, XP_308607.3)
Lethal (Dme, NP_523927.1)
protein lethal (Phu, EEB15480.1)
similar to Protein Ett21 (Ame, XP_394333.2)
similar to Hsp21.7 (Nvi, XP_001604985.1)
similar to ENSANGP00000018891 (Nvi, XP_001605148.1)
similar to small Hsp21 (Tca, XP_974390.1)
similar to small Hsp21 iso1 (Tca, XP_973442.1)
similar to Hsp1 (Nvi, XP_001600020.1)
similar to Protein Ett21 (Ame, XP_001119830.1)
similar to ENSANGP00000018891 (Nvi, XP_001604562.1)
similar to small Hsp21 (Tca, XP_966780.1)
similar to small Hsp21 iso1 (Tca, XP_974390.1)
similar to Hsp21 (Nvi, XP_001604469.1)
hypothetical protein C43 (Cel, NP_509045.1)
hypothetical protein (Cin, XP_002119131.1)
Hsp 12.2 (Cel, NP_498776.1)
(Bom, BGIBMGA004101)
Hsp19.9 (Bom, NP_001036984.1)
similar to small Hsp21 (Tca, XP_966780.1)
similar to small Hsp21 iso1 (Tca, XP_974390.1)
similar to ENSANGP00000018891 (Nvi, XP_001605148.1)
similar to Hsp21 7 (Nvi, XP_001604985.1)
similar to Protein Ett21 (Ame, XP_001119830.1)
similar to Hsp1 (Nvi, XP_001600020.1)
similar to Protein Ett21 (Ame, XP_394333.2)
similar to Lethal (Dme, NP_523927.1)
AGAP007161 (Aga, XP_308607.3)
AGAP007162 (Aga, XP_308606.2)
(Bom, BGIBMGA008574)
Hsp70 (Dpu, NCBI_GNO_322413)
hypothetical protein C30C11.4 (Cel, NP_498868.1)
Hsp70 (Dpu, NCBI_GNO_370774)
Hsp70Ba (Tca, XP_973521.1)
Hsc2 (Dme, NP_524339.1)
similar to Hsp70 (Nvi, XP_001599773.1)
similar to Hsp70 (Nvi, XP_001037396.1)
Hsp70 (Bom, NP_001013589.1)
Hsp70-2a (Phu, EEB14908.1)
(Tca, GLEAN_05094)
Hsp70-2 (Phu, EEB14951.1)
similar to Hsp68 (Tca, XP_001814746.1)
Hsp70Aa (Tca, XP_974442.1)
similar to Hsp70 (Tca, XP_971651.1)
Hsp68 (Dme, NP_524414.1)
Hsp70Aa (Dme, NP_731651.1)
Hsp70Ba (Dme, NP_731716.1)
Hsp70Bb (Dme, NP_524921.2)
Hsp70Bbb (Dme, NP_788663.1)
similar to Hsp70 (Tca, XP_001811933.1)
Hsp1 (Cel, NP_503088.1)
similar to Hsc70 (Cin, XP_002129101.1)
(Dpu, NCBI_GNO_306193)
(Dpu, NCBI_GNO_40593)
similar to Hsc70 (Nvi, XP_001608044.1)
similar to Hsp8 iso1 (Ame, XP_623130.1)
Hsc1 isoA (Dme, NP_524063.1)
AGAP004944 (Aga, XP_315042.3)
GG16900 (Dme, XP_001980016.1)
similar to Hsp8 iso 1 (Ame, XP_392933.2)
similar to Hsc70 (Nvi, XP_001607939.1)
Hsc70 (Phu, EEB10106.1)
Hsc (Bom, NP_001036892.1)
similar to Hsc70 iso1 (Tca, XP_966611.1)
similar to CG7033 (Ame, XP_393300.2)
similar to ENSANGP00000004677 (Nvi, XP_001602878.1)
AGAP009694 (Aga, XP_318752.4)
(Dpu, NCBI_GNO_584363)
(Dpu, NCBI_GNO_22633)
(Phu, EEB16674.1)
T-complex protein 1 subunit beta (Phu, EEB16405.1)
similar to t-complex polypeptide (Cin, XP_002119405.1)
Chaperonin Containing TCP-1 family member (Cel, NP_741153.1)
chaperonin subunit 6a zeta (Bom, NP_001040108.1)
similar to chaperonin subunit 6a zeta (Tca, XP_967748.1)
XP_312164.4 (Aga, XP_312164.4)
similar to TCP-1-gamma (Ame, XP_392814.2)
Hsp60D (Dme, NP_609656.1)
Hsp60 related (Dme, NP_524925.1)
(Cin, ENSCINP00000016023)
Hsp60 (Cel, NP_497429.1)
Hsp60C (Dme, NP_608848.2)
Hsp60 (Tca, XP_971630.1)
similar to Hsp60 (Tca, XP_971630.1)
similar to Hsp60 (Ame, XP_392899.3)
(Bom, BGIBMGA007349)
similar to ENSANGP00000014839 (Nvi, XP_001600045.2)
Hsp60 isoA (Dme, NP_511115.2)
AGAP004002 (Aga, XP_318461.2)
Hsp60 (Phu, EEB15946.1)
(Dpu, NCBI_GNO_542253)
(Bom, BGIBMGA005780)
hypothetical protein Y55F3BR.6 (Cel, NP_499998.3)
AGAP000941 (Aga, XP_560153.3)
Hsp25 (Cel, NP_001024374.1)
CG14207 isoB (Dme, NP_728275.1)
similar to Hsp20.6 iso3 (Nvi, XP_001607678.1)
similar to Hsp20.6 iso2 (Nvi, XP_001607675.1)
similar to Hsp20.6 iso1 (Nvi, XP_001607669.1)
similar to Hsp20.6 (Tca, XP_973885.1)
similar to CG14207 (Ame, XP_392405.3)
Hsp21.4 (Bom, NP_001036983.1)
Hsp21.4 (Api, ACYPI003907)
Hsp20.6 (Dpu, NCBI_GNO_782103)
(Bom, BGIBMGA009823)

Figure S2. **Alignments of putative antimicrobial peptides megourin and penaeidin. (**a) Putative pea aphid megourin (pea_aphid) aligned with 3 megourins of the aphid *Megoura viciae* (MEGVI). (b) Putative pea aphid penaeidin (ACYPI37769-PA) aligned with penaeidins from the shrimp *Penaeus vannamei* (PENVA) and *Penaeus (Litopenaeus) setiferus* (LITS).

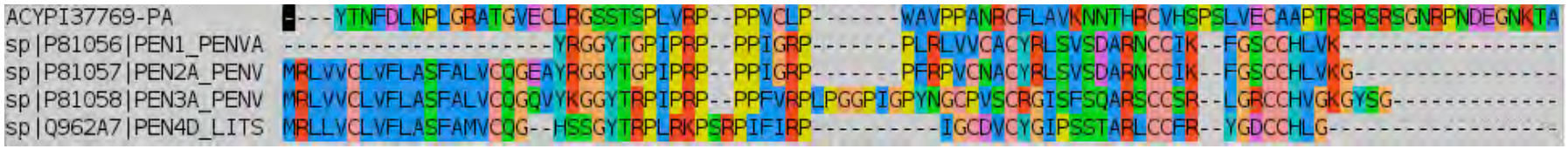
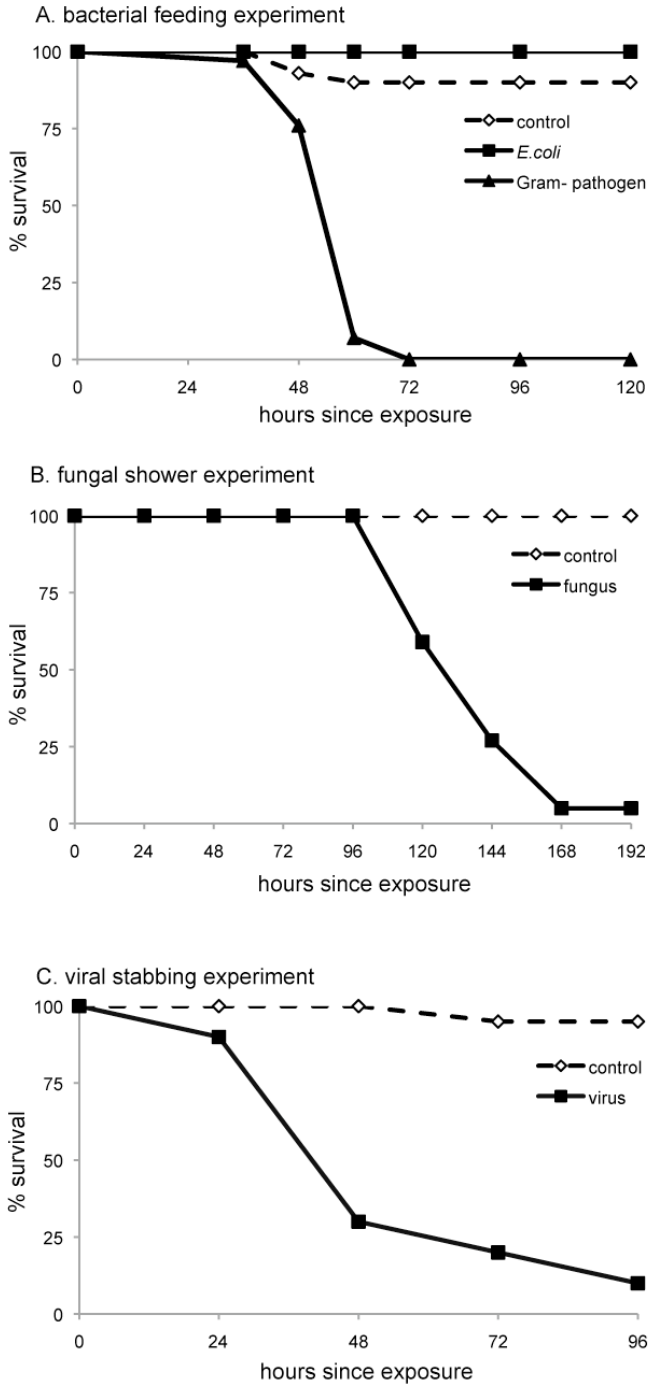


Figure S3. Survival curves for experimental infections associated with the qPCR study. (a) In the bacterial feeding experiment, 30 aphids per condition were fed for 20hrs on artificial diet containing bacteria or on control diet and then transferred to plants and monitored for survival. Though survival data for the bacterial stabbing experiment, in which aphids were stabbed with bacteria-contaminated needles, was not collected daily, after two days, overall survival was similar to the feeding trials, with 100% survival of control aphids, 80% survival of *E. coli*-stabbed aphids, 30% survival of Gram-positive pathogen-stabbed aphids, and 0% survival of Gram-negative pathogen-stabbed aphids (n = 10 per condition). (b) In the fungal shower experiment, 22 aphids were exposed to a shower of fungal spores for two hours and then transferred to plants. Control aphids were not exposed. (c) In the viral stabbing experiment, 20 aphids were stabbed with a virus-contaminated needle and then transferred to plants. Control aphids were not stabbed.



1. Koga R, Tsuchida T, Sakurai M, Fukatsu T: Selective elimination of aphid endosymbionts: effects of antibiotic dose and host genotype, and fitness consequences. *Fems Microbiology Ecology* 2007, 60: 229-239.
2. Harada H, Ishikawa H: Experimental pathogenicity of *Erwinia aphidicola* to pea aphid, *Acyrthosiphon pisum*. *Journal of General and Applied Microbiology* 1997, 43: 363-367.
3. Rahbe Y, Febvay G: Protein toxicity to aphids - an invitro test on *Acyrthosiphon pisum*. *Entomologia Experimentalis Et Applicata* 1993, 67: 149-160.
4. Mietkiewski R, Soper RS, Balazy S: Notes on *Zoophthora occidentalis* (Thaxter) Batko (entomophthorales, entomophthoraceae) *Mycotaxon* 1981, 13: 41-49.
5. Williamson C, Rybicki EP, Kasdorf GGF, Vonwechmar MB: Characterization of a new picorna-like virus isolated from aphids. *Journal of General Virology* 1988, 69: 787-795.
6. van Munster M, Dullemans AM, Verbeek M, van den Heuvel J, Clerivet A, van der Wilk F: Sequence analysis and genomic organization of Aphid lethal paralysis virus: a new member of the family Dicistroviridae. *Journal of General Virology* 2002, 83: 3131-3138.
7. Wilson ACC, Dunbar HE, Davis GK, Hunter WB, Stern DL, Moran NA: A dual-genome microarray for the pea aphid, *Acyrthosiphon pisum*, and its obligate bacterial symbiont, *Buchnera aphidicola*. *Bmc Genomics* 2006, 7.
8. Rozen S, Skaletsky HJ: Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Edited by Krawetz S, Misener S. Totowa, N.J.: Humana Press; 2000: 365-386.
9. Pea Aphid Gbrowse [http://genoweb1.irisa.fr/cgi-bin/gbrowse/gbrowse/aphidbase/]
10. AphidBase Downloads [http://www.aphidbase.com/aphidbase/downloads]
11. *Acyrthosiphon pisum* EST Database [http://aphidests.org/]